

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:24:24 ; Search time 25.25 seconds
(without alignments)
290.327 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTRLVTAALLGLLMVV.....APRRASEPKHKNQAEIAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235.5	19.2	187	1	PEBP_RAT
2	234.5	19.2	221	1	PBPH_CAEEL
3	231	18.9	152	1	D1_ONCVO
4	223	18.2	134	1	D3_ONCVO
5	222	18.1	197	1	OV16_ONCVO
6	219	17.9	114	1	D2_ONCVO
7	219	17.9	187	1	PEBP_MOUSE
8	218.5	17.9	186	1	PEBP_BOVIN
9	206	16.8	186	1	PEBP_MACFA
10	205	16.7	186	1	PEBP_HUMAN
11	169.5	13.8	262	1	TE26_TOXCA
12	168	13.7	210	1	OBA5_DROME
13	148	12.1	181	1	CEN_ANTMA
14	137.5	11.2	190	1	PBP_PLAFA
15	133	10.9	201	1	YL79_YEAST
16	122.5	10.0	219	1	DKAL_YEAST
17	113.5	9.3	157	1	YJ10_MYCTU
18	113	9.2	151	1	Y273_METH
19	106	8.7	171	1	YC50_AQUAE
20	105	8.6	198	1	YC69_PYRHO
21	96.5	7.9	150	1	Y877_CHLPN
22	95	7.8	183	1	YBCL_ECOLI
23	95	7.8	201	1	YJ11_MYCTU
24	92	7.5	179	1	Y502_SIRCO
25	91.5	7.5	551	1	SYE_ARCFU
26	89.5	7.3	176	1	YL40_MYCTU
27	89.5	7.3	1705	1	PTPO_MOUSE
28	86.5	7.1	150	1	Y736_CHLTR
29	85	6.9	595	1	SNX3_HUMAN
30	80.5	6.6	287	1	Y98_ARCFU
31	79	6.5	265	1	HL_PEA
32	79	6.5	602	1	TX15_MOUSE
33	77.5	6.3	490	1	ACM4_CHICK

34	76	6.2	748	1	MEPA_RAT
35	76	6.2	1102	1	PL1G_PIG
36	76	6.2	2607	1	BACB_BACLI
37	75.5	6.2	1130	1	REPT_MOUSE
38	75	6.1	466	1	ACM2_CHICK
39	75	6.1	1041	1	EGT2_YEAST
40	74.5	6.1	630	1	INLB_LISMO
41	74	6.0	659	1	GLGX_HAEIN
42	74	6.0	1004	1	YD83_SCHPO
43	73.5	6.0	240	1	GTXH_MOUSE
44	73.5	6.0	382	1	YMF4_CAEEL
45	73.5	6.0	397	1	TBXT_CHICK

ALIGNMENTS

RESULT	PEBP_RAT	STANDARD	PRT	187 AA
ID	PEBP_RAT	STANDARD	PRT	187 AA
AC	P31044; P31045			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN (PEBP) (23 KDA MORPHINE-BINDING PROTEIN) (P23K).			
DE	BINDING PROTEIN (P23K).			
GN	PEBP OR PBP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
[1]	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;			
RX	MEDLINE=91042640; PubMed=1978248;			
RA	Grandy D.K., Hanneman E., Bunzow J., Shih M., Machida C.A., Bidlack J.M., Civelli O.			
RT	"Purification, cloning, and tissue distribution of a 23-kDa rat protein isolated by morphine affinity chromatography.";			
RL	Mol. Endocrinol. 4:1370-1376(1990).			
[2]	SEQUENCE FROM N.A.			
RC	STRAIN=WISTAR; TISSUE=HIPPOCAMPUS;			
RA	Tohdoh N., Tojo S., Agui H., Ojika K.;			
RN	Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.			
[3]	SEQUENCE FROM N.A.			
RC	STRAIN=WISTAR; TISSUE=EPIDIDYMIS, AND LIVER;			
RX	MEDLINE=94311839; PubMed=8037677;			
RA	Perry A.C.F., Hall L., Bell A.E., Jones R.;			
RT	"Sequence analysis of a mammalian phospholipid-binding protein from testis and epididymis and its distribution between spermatozoa and extracellular secretions.";			
RL	Biochem. J. 301:235-242(1994).			
[4]	SEQUENCE OF 40-56 AND 93-112.			
RC	TISSUE=SPERM;			
RX	MEDLINE=92031654; PubMed=1932083;			
RA	Jones R., Hall L.;			
RT	"A 23 kDa protein from rat sperm plasma membranes shows sequence similarity and phospholipid binding properties to a bovine brain cytosolic protein.";			
RL	Biochim. Biophys. Acta 1080:78-82(1991).			
CC	-I- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE. HAS LOWER AFFINITY FOR PHOSPHATIDYLINOSITOL AND PHOSPHATIDYLCHOLINE.			
CC	-I- SUBUNIT: HAS A TENDENCY TO FORM DIMERS BY DISULFIDE CROSS-LINKING.			
CC	-I- SUBCELLULAR LOCATION: THERE ARE TWO FORMS: A CYTOPLASMIC FORM AND A MEMBRANE-BOUND FORM.			
CC	-I- TISSUE SPECIFICITY: MAJOR COMPONENT OF EPIDIDYMAL SECRETIONS AND SPERM PLASMA MEMBRANES. IT IS PRESENT IN CYTOSOLS FROM A VARIETY OF OTHER TISSUES.			
CC	-I- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/send_an_email_to_license@sib-sib.ch).

DR	EMBL; X75253; CAA53032.1; -	
DR	EMBL; X75254; CAA53033.1; -	
DR	EMBL; X71873; CAA50708.1; -	
DR	PIR; A36126; A36126.	
DR	PIR; S37555; S37555.	
DR	PIR; S18358; S18358.	
DR	INTERPRO; IPR001838; -	
DR	PFAM; PF01161; PBP; 1.	
DR	PROSITE; PS01220; PBP; 1.	
KW	Lipid-binding.	?
FT	MOD_RES	?
FT	CONFLICT	49
FT	CONFLICT	54
FT	CONFLICT	55
SO	SEQUENCE	187 AA; 20801 MW; F2BF053FE34B8056 CRC64; F2BF053FE34B8056 CRC64;

	Query Match	19.2%	Score 235.5;	DB 1;	Length 187;
	Best Local Similarity	36.5%;	Pred. No. 1.2e-15;		
	Matches	57;	Conservative	20;	Mismatches 60;
				Indels	19; Gaps
Qy	52	ELGNIGKVVDPDCNNYRQKITTSWMEPIVKFPGVGDGATYILVMVDPAAPRAEPRQRFWR	111		
Dd	36	ELGKV---LTPTQVMNRPSSISW-----DGLDFGKLYTLVLTPDAPSRRDKPFREWH	85		
Qy	112	HMLVTDIKGADLKKGGIKQGELSAYQAOPSPAHSGFHRYQFYQLQEGKVIS----	LLPK 167		
Dd	86	HFLVNMKNNDISSGTV----LSEYVGSGPKDGLHRYVWLVEQEQLNCDEPILSNK	141		
Qy	168	ENKTRGSWMKDRLFNRPHLGEPFEASTQFWMTQNYODS	203		
Dd	142	SGDNRGKPKVESFRKKYHLGAPVAGTCFOAE-WDDDS	176		

```

RESULT 2
PBPH_CAEEL
ID PBPH_CAEEL STANDARD; PRT; 221 AA.
AC O16264;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN HOMOLOG F40A3.3.
GN F40A3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Bradshaw H., Keppler D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>,
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF016423; AAB65322.1; --
 DR WORMPEP: F40A3.3; CEI0146.
 DR INTERPRO: IPR001858; --
 DR PFAM: PF01161; PBP; 1.

```

DR PROSITE: PS01220; pep: 1.
KW Hypothetical protein; Lipid-binding.
SQ SEQUENCE 221 AA; 24143 MW; 4828CC749DB832AC CRC64;

Query Match 19.2%; Score 234.5; DB 1; Length 221;
Best Local Similarity 33.1%; Pred. No. 1.8e-15;
Matches 58; Conservative 16; Mismatches 60; Indels 41; Gaps
QY 59 KVVPP--CNYRQKITSNMETPVKPPGAVD-----GATYIL 92
Db 46 EVLPDVLASNPSSKVVS-----VRENSGVANLGNLTPQVKDTPVKWDAEPGALYTL 10
QY 93 VMYDDPAPSGRAEPRFRWHLVTDIKGADLKKGKIQOELSAYQSPSPAHSGFHYQF 15
Db 101 IKTDPPAPSRKEPTREWHHLVWVNPIGNDIK----GDTLSEYIGAGPPKPTGLHRYVI 15
QY 153 FVYLGQSKVIS-----LLPKENKTRGSKMKDRFLNRFHGLPEASTQFMPTNQD 202
Db 157 YVYQCGEPTDPAHCEPLTWTSCKRGEGWKAADFVAKHKIGAPVFGNLFQAE-YDD 210

```

```

RESULT 3
DI_ONCYO
ID DI_ONCYO STANDARD; PRT; 152 AA.
AC P54186;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DE DI PROTEIN (FRAGMENT).
GN DI.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97045813; PubMed=8890735;
RA Ertmann K.D., Gallin M.Y.;
RT "Onchocerca volvulus: Identification of cDNAs encoding a putative
RT phosphatidyl-ethanolamine-binding protein and a putative partially
RT processed mRNA precursor.";
RL Gene 174:203-207(1996).

```

PROTEIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC      -----
DR      EMBL; X87991; CRA61244.1; -.
DR      INTERPRO; IPR001858; -.
DR      PFAM; PF01161; PBP; 1.
DR      PROSITE; PS01220; PBP; 1.
DR      NON_TER 1
FT      SEQUENCE 152 AA; 16767 MW; E3B68C2E3E3E295D CRC64;
SO

```

	Query Match	18.9%	Score 231;	DB 1;	Length 152;
	Best Local Similarity	37.0%;	Pred. No. 2.5e-15;		
	Matches 54;	Conservative 18;	Mismatches 58;	Indels 16;	Gaps 5;
QY	55	NIGCKVVP-DCMNYRQKTTSMMEPIVKFPGAVDGGATYILVMVDDPDAPSRAEPRQRFNRHW	113		
Db	20	NLGNLTPTQVKNQPTKV-SW-----DAEPGALYTLVMTDPDAPSKRKNPVFRENHHW	70		
QY	114	LVYTDKGLDKKGIQGGELSAYQAPSPPAHSGFHRQYFFVYVLOQSKYISLLLPENKTRG	173		
Db	71	YINISGCVNVSSTGV-----ISDYICSGPKPGLHRYFLVYVKKPGSLTD--TOHGGNRP	124		

[illegible][illegible]

KW	Lipid-binding.
SO	SEQUENCE 187 AA; 20860 MW; E51DD693A0A8BABD CRC64;

Db 24 LQVKYGGAEVDELGKVLTPQVKNRPTSIW-----DGLDPGKLYTLVLTPDPAPSrk 76

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D16111; BAA03684.1; -

CC EMBL; X72522; CAA53031.1; -

CC EMBL; X85033; CAA59404.1; -

CC EMBL; S76773; RAD14234.1; -

CC PDB; 1BD9; 16-SEP-98.

CC PDB; 1BEH; 16-SEP-98.

CC SWISS-2DPAGE; P30086; HUMAN.

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

CC PROSITE; PS01220; PBP; 1.

CC Lipid-binding; 3D-structure.

CC INIT_MET 0 0

CC VARIANT 8 8

CC S -> N.

CC CONFLICT 7 7 /FTID=VAR_006048.

CC W -> K (IN REF. 2).

CC SEQUENCE 186 AA; 20925 MW; F18F2AC30747C78B CRC64;

Query Match 16.7%; Score 205; DB 1; Length 186;

Best Local Similarity 33.1%; Pred. No. 1e-12;

Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNTGCKVPCNNYRQKITSWMEPIVKFPGAVDGYIILVWYDPDAPSPRAEPQRFWR 111

DB 35 ELGKV---LTPTQVKNRPTISW-----DGLSDGKGLYTLVLDTPDAPSRKDPKYREWH 84

QY 112 HWLVTDIKGADLKKGIQQLSAYQAPSPPAHSGFHRQYFVYVQLQEGKVIS-----LLPK 167

DB 85 HFLVVMKGNLSSGTV-----LSDYVGSPPKGTGLHRYVWLVEYQDRPLKCDPEILSNR 140

QY 168 ENKTRGSKMDRFLNRFLHGEPEASTQPMQTQ 198

DB 141 SCDHRGKFKVASFRRKKYELRPAVAGTCYQAE 171

RESULT 11

TE26_TOXCA STANDARD; PRT; 262 AA.

ID TE26_TOXCA STANDARD; PRT; 262 AA.

AC P54190;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 26 KDA SECRETED ANTIGEN PRECURSOR (TOXOCARA EXCRETORY-SECRETORY ANTIGEN-26) (TES-26).

OS Toxocara canis.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;

OC Toxocaridae; Toxocara.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95355481; PubMed=7629180;

RA Gems D., Ferguson C.J., Robertson B.D., Nieves R., Page A.P., Blaxter M.L., Maizels R.M.

RT "An abundant, trans-spliced mRNA from Toxocara canis infective larvae encodes a 26-kDa protein with homology to phosphatidylethanolamine-binding proteins."

RL J. Biol. Chem. 270:18517-18522(1995).

CC -1- FUNCTION: BINDS PHOSPHATIDYLETHANOLAMINE.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- DEVELOPMENTAL STAGE: FOUND IN LARVA, BUT NOT IN THE ADULT PARASITE.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

CC EMBL; U29761; AAC46843.1; -

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

CC PROSITE; PS01220; PBP; 1.

CC Lipid-binding; Antigen; Signal.

CC SIGNAL 1 21 POTENTIAL.

CC FT CHAIN 22 262 26 KDA SECRETED ANTIGEN.

CC SQ SEQUENCE 262 AA; 28048 MW; 426ED0DESCC070AL CRC64;

Query Match 13.8%; Score 169.5; DB 1; Length 262;

Best Local Similarity 27.2%; Pred. No. 4e-09;

Matches 56; Conservative 27; Mismatches 86; Indels 37; Gaps 5;

QY 25 DENSPCAHEALLDEDTLF-----COGLEVPYPELGNIGCKVVP----- 62

DB 61 DEANCAASINLCQNPTEFPLVRDRQCQKTCGLCAGCGFSSGIVPLVVTSPSRVSVTF 120

QY 63 -----DCNNYRQKITSWMEPIVKFPGAVDGYIILVWYDPDAPSPRAEPQRFWRHVLVT 116

DB 121 ANNQVNCNGTTLTAOVAQPTVWE-AQPNDRYTLIMVDPDPSPSAAGQOQORLHWWI 179

QY 117 DIKGADLKKGIQQLSAYQAPSPPAHSGFHRQYFVYVQLQEGKV-----ISLLPKENKTR 172

DB 180 NIPGNNI-----AGGTTAAFPQSTPAANTGVHRYVELVYRQPAAINSLNLLNVQDSR 235

QY 173 GSWKMDRFLNRFLHGEPEASTQPMQTQ 198

DB 236 PGFGTTAFATQFNLGSPYAGNFYRSQ 261

RESULT 12

OBA5_DROME STANDARD; PRT; 210 AA.

ID OBA5_DROME STANDARD; PRT; 210 AA.

AC P54185;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE PUTATIVE ODORANT-BINDING PROTEIN A5 PRECURSOR (ANTENNAL PROTEIN 5).

GN A5.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CANTON-S; TISSUE=ANTENNA;

RX MEDLINE=94121915; PubMed=7545907;

RA Pikielny C.W., Hasan G., Rouyer F., Rosbash M.

RT "Members of a family of Drosophila putative odorant-binding proteins are expressed in different subsets of olfactory hairs."

RL Neuron 12:35-49(1994).

CC -1- SUBCELLULAR LOCATION: SECRETED IN THE LUMEN OF THE SENSILLA (POTENTIAL).

CC -1- TISSUE SPECIFICITY: CELLS AT THE BASES OF A FEW SCATTERED SENSILLA ON THE POSTERIOR SURFACE OF THE ANTENNA.

CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U05243; AAC46472.1; -

CC FLYBASE; FBgn0011294; a5.

CC INTERPRO; IPR001858; -

CC PFAM; PF01161; PBP; 1.

```

KW 3D-structure.
SQ SEQUENCE 181 AA; 20322 MW; 32673415FE29E503 CRC64;

Query Match                12.1%; Score 148; DB 1; Length 181;
Best Local Similarity      29.1%; Pred. No. 2.9e-07;
Matches 44; Conservative 25; Mismatches 62; Indels 20; Gaps 5;

QY 60 VVPDCNN-----YRKITSMPEIVKFFPGAVDGATYIILVMVDPDAPSRAEPQR 108
   : : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 31 VIYNNSNISKHYVNGHELFPASVTS--TPRVEHGDWMSFFTLLMTDPPVGPSDPLYR 88

QY 109 FWRHLVTDIKG-ADLKKGIQOELSAYQAPSPAINSGCFHRFYOFFVYLQGGKVISLLPK 167
   ||| |||| | | | | : | | : | | : | | : | | : | | : | |
Db 89 EHLHWIVIDIPGTDDSSFGK----EVVSYEMLPRP--NIGIHFRVFLLFKOKKGQAMLSP 142

QY 168 ENKTRGSWKMDRFLNRFLHGLGEPEASTOFMTQ 198
   : : | | : | | : | | : | | : | | : | | : | | : | |
Db 143 PIVCRDGFNTKRKTQENELGPLVAAVFENCQ 173

RESULT 14
PBP_PLAFA
ID PBP_PLAFA STANDARD; PRT; 190 AA.
AC P54189;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PUTATIVE PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP MEDLINE=95364848; PubMed=7637711;
RX Trottein F., Cowman A.F.;
RA "The primary structure of a putative phosphatidylethanolamine-binding
RT protein from Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 70:235-239(1995).
CC -! SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18984; AAC47026.1; -.
CC DR INTERPRO; IPR001858; -.
CC DR PFAM; PF01161; PBP; 1.
CC DR PROSITE; PS01220; PBP; 1.
CC KW Lipid-binding.
CC SQ SEQUENCE 190 AA; 21576 MW; 1EA279A1C81ICE41 CRC64;

Query Match                11.2%; Score 137.5; DB 1; Length 190;
Best Local Similarity      31.5%; Pred. No. 3.2e-06;
Matches 35; Conservative 20; Mismatches 43; Indels 13; Gaps 3;

QY 86 DGATYIILVMVDPAAPSRAEPRORFWRHVLVTDIKDGLKKGIQG-QELSAYQAPSPAHP 144
   || : | | | | | | | | | | : | | : | | : | | : | | : | |
Db 66 DGVCFVLFMWDPDPSRLRPDCKEYIIHWWSGKITKELIKGTQNCVTLIPVGPSIKKG 125

QY 145 SGFHRYOFFVYLQGGKVISLLPKENKTR---GSWKMDRFLNRFLHGLGEPA 191
   : | | | : | | : | | : | | : | | : | | : | | : | |
Db 126 TGLHRISF-----IISLKEDKNITGLPHYKGEKYITRVKFNNYES 168

RESULT 15
YL79_YEAST
CD YL79 YEAST STANDARD; PRT; 201 AA.
```

Sat Mar 24 13:27:34 2001

```

AC Q06252;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOPHETICAL 22.2 KDA PROTEIN IN TFS1-SAM1 INTERGENIC REGION.
GN YLR179C OR L9470.20.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifkin L., Riles L., Raich A., Trevaskis E., Vignati D., R.;
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17246; AAB67472.1; -
CC SGD; S0004169; YLR179C.
CC INTERPRO; IPR001858; -
CC PFAM; PF01161; PBP; 1.
CC PROSITE; PS01220; PBP; 1.
CC Hypothetical protein
KW SEQUENCE 201 AA; 22151 MW; 00157450B02451A2 CRC64;
SQ

```

Query Match 10.9%; Score 133; DB 1; Length 201;
Best Local Similarity 31.0%; Pred. No. 9.2e-06;
Matches 39; Conservative 19; Mismatches 38; Indels 30; Gaps 6;

```

QY 92 LVMVDPDAPSRAEPRQRWRHVLVTDI-----KGADLK---KGKIQGQELSAYQAPSPPA 143
DB 75 LLMTDPDAPSRTEHKWSEVCHYIITDIPVEYCGDGTIAISGKGVVRNN-----YIGPGPPK 130
QY 144 HSGPHRYOFFVYVLOEGKVISLLPKENKTRGSKMDRFLNRPHLGEP-----EASTQ 194
DB 131 NSGYHYRYVEFLCKQ-----PKGADSTFTKVENTIIS-WGYGTPGAGAYDIKENNLQ 181
QY 195 FMTQNY 200
DB 182 LVGANV 187

```

Search completed: March 24, 2001, 13:24:26
Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:28 ; Search time 47.63 Seconds
(without alignments)
162,964 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTRLVTAALLGLMMVV.....APRRASEPKHKNOAIAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1224	100.0	227	20	Y35976 Extended human sec
2	1224	100.0	227	21	Y64647 Human phosphatidyl
3	1213	99.1	227	21	Y94263 Human phospholipid
4	652	53.3	121	20	Y11860 Human 5' EST seque
5	235.5	19.2	187	13	R27897 HCNP protein. Rat
6	235.5	19.2	187	15	R49942 Rat hippocampal ch
7	231	18.9	152	21	Y94265 Onchocerca volvulu
8	205	16.7	187	13	R27718 HCNP precursor pro
9	205	16.7	187	15	R49943 Human hippocampal
10	205	16.7	187	16	R64268 Phosphatidylethano
11	164	13.4	175	20	Y49098 Amino acid sequenc
12	148	12.1	181	18	W13944 Antirrhinum centro

13	139	11.4	152	12	R15223	OV-16 antigen. On
14	137.5	11.2	177	18	W13945	Arabidopsis termin
15	89.5	7.3	120	20	Y11503	Human 5' EST seque
16	89.5	7.3	176	19	W72901	Mycobacterium tube
17	89.5	7.3	176	20	Y21918	Amino acid sequenc
18	86.5	7.1	156	20	Y37418	Amino acid sequenc
19	81	6.6	234	12	R13050	CD4-specific CDR-g
20	79.5	6.5	232	16	R80616	Anti-human IL-4 hu
21	77	6.3	110	20	Y35525	Chlamydia pneumoni
22	77	6.3	579	20	Y14130	Human IL-1R18 prot
23	77	6.3	875	19	W37098	Murine endothelial
24	76.5	6.2	236	21	Y56286	HCAT1 clone 25 ant
25	76.5	6.2	321	21	Y92341	Human cancer assoc
26	76.5	6.2	490	21	Y56637	HCAT1 binding huma
27	76	6.2	233	14	R30777	pH52-9.0 humanise
28	75	6.1	585	20	W97757	S-region transfer
29	75	6.1	670	20	Y34517	Porphyromonas ging
30	75	6.1	702	20	Y34390	Porphyromonas ging
31	74.5	6.1	658	20	Y21164	Human TIGIR prote
32	74.5	6.1	686	20	Y14128	Human IL-1R18 prot
33	74	6.0	548	20	Y06927	C. albicans antige
34	73.5	6.0	1192	20	Y23899	Human resenilin bi
35	73.5	6.0	1211	18	W24560	Presenilin-interac
36	73	6.0	458	21	Y81990	Chimeric PCB decom
37	73	6.0	1049	17	W11576	Human phosphatidyl
38	73	6.0	1050	17	W11577	Human phosphatidyl
39	73	6.0	1101	19	W23947	Human phosphoinosi
40	73	6.0	1101	20	W90089	Human G-protein re
41	73	6.0	1101	21	Y78803	Human p120 regulat
42	73	6.0	1711	19	W70506	Osteostesticular pr
43	73	6.0	1711	19	W70507	Mutant osteostestic
44	72.5	5.9	327	21	Y83092	F-box protein FBP-
45	72.5	5.9	503	19	W49909	Signal regulatory

ALIGNMENTS

RESULT 1

Y35976
ID Y35976 standard; Protein; 227 AA.

XX Y35976;

XX AC Y35976;

XX DT 13-SEP-1999 (first entry)

XX DE Extended human secreted protein sequence, SEQ ID NO. 225.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW genetic disease.

XX OS Homo sapiens.

XX PN WO9931236-A2.

XX PD 24-JUN-1999.

XX PF 17-DEC-1998; 98WO-IB02122.

XX PR 10-AUG-1998; 98US-0096116.

XX PR 17-DEC-1997; 97US-0069957.

XX PR 09-FEB-1998; 98US-0074121.

XX PR 13-APR-1998; 98US-0081563.

XX PA (GEST) GENSET.

XX PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX WP1; 1999-385906/32.

XX DR N-PSDB; X97660.

xx PT New isolated human secreted proteins
xx PS Claim 9; Page 255; 516pp; English.
xx CC This sequence is encoded by an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
xx SQ Sequence 227 AA;

Query Match 100.0%; Score 1224; DB 20; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.9e-131;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWMTMLVTAALLGLMMVVTGDEDENSPCAHEALDEDTLFCQGLEVFPELGNIGCKV 60
DB 1 mgwtmrlvtlaallglmmvvtgdedenspcachealldeedtlfcqglevfpeignigckv 60

QY 61 VPCNNYRQKITSMWEPVVKFPGAVDGTATLVMVDPDAPSPRAEPFRFWRHLVTDIKG 120
DB 61 vpcdcnyrqkkitwmeipivkfpavdgatylvmvdpdapsraepqrgrfwrhlvtdikg 120

QY 121 ADLKKGIQGGELSAQAPSPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRGSKWMDRF 180
DB 121 adlkkgkigqgelsaqapsppahsgfhrfyqffvylqegkvisllpkentkrgswkmdrf 180

QY 181 LNRPHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKHKNQAEIAAC 227
DB 181 lnrphlgepeastqfntqnyqdsptlqapreraspkhknqaeiaac 227

RESULT 2
Y64647
ID Y64647 standard; Protein: 227 AA.
XX AC Y64647;
XX DT 01-FEB-2000 (first entry)
DE Human phosphatidylethanolamine-binding protein.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification.
XX OS Homo sapiens.
XX PN WO9953051-A2.
XX PD 21-OCT-1999.
XX PF 09-APR-1999; 99WO-IB00712.
XX PR 09-APR-1998; 98US-0057719.
XX PR 28-APR-1998; 98US-0069047.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX

DR WPI: 2000-038446/03.
DR N-PSDB; 242252.
XX Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX Example 21; Page 169-170; 837pp; English.
XX 242265 to 243075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. Y64651 to Y64653
CC represent the EST-related proteins corresponding to 242265 to 243052.
CC The 5' ESTs can be used for producing secreted human gene products.
CC They can be used to identify and isolate 5' untranslated regions (UTRs)
CC and upstream regulatory regions which control the location, development
CC stage, rate, and quantity of protein synthesis, as well as stability of
CC mRNA. The ESTs are also useful as probes for chromosome mapping, and to
CC obtain full length cDNA clones. The ESTs can also be used in forensic
CC procedures to identify individuals, or in diagnostic procedures to
CC identify individuals having genetic diseases resulting from abnormal
CC gene expression. The products may also be used in gene therapy protocols.
CC The nucleic acids encoding signal peptides can be used for directing
CC extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
CC The proteins encoded by the EST sequences may be useful in treating a
CC variety of human conditions. Secreted proteins have therapeutic value,
CC and the identification of new secreted proteins is valuable. 242249 to
CC 242264 and Y64644 to Y64650 represent sequences used in the
CC exemplification of the present invention.
xx SQ Sequence 227 AA;

Query Match 100.0%; Score 1224; DB 21; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.9e-131;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWMTMLVTAALLGLMMVVTGDEDENSPCAHEALDEDTLFCQGLEVFPELGNIGCKV 60
DB 1 mgwtmrlvtlaallglmmvvtgdedenspcachealldeedtlfcqglevfpeignigckv 60

QY 61 VPCNNYRQKITSMWEPVVKFPGAVDGTATLVMVDPDAPSPRAEPFRFWRHLVTDIKG 120
DB 61 vpcdcnyrqkkitwmeipivkfpavdgatylvmvdpdapsraepqrgrfwrhlvtdikg 120

QY 121 ADLKKGIQGGELSAQAPSPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRGSKWMDRF 180
DB 121 adlkkgkigqgelsaqapsppahsgfhrfyqffvylqegkvisllpkentkrgswkmdrf 180

QY 181 LNRPHLGEPEASTQFMTQNYQDSPTLQAPRERASEPKHKNQAEIAAC 227
DB 181 lnrphlgepeastqfntqnyqdsptlqapreraspkhknqaeiaac 227

RESULT 3
Y94263
ID Y94263 standard; protein; 227 AA.
XX AC Y94263;
XX DT 01-AUG-2000 (first entry)
XX DE Human phospholipid binding protein 2, PLBP2.
XX KW Human; phospholipid binding protein; PLBP2; foetal development disorder;
KW reproduction disorder; cell proliferation disorder; immune response;
KW autoimmune disorder; AIDS; infertility; cytostatic; immunosuppressive;
KW gene therapy; hereditary neuropathy;
KW phosphatidylethanolamine binding protein D1; PE-BP D1.
XX OS Homo sapiens.
XX PF key
XX FT Location/Qualifiers
FT Peptide 1..21

Db 36 elgkv---ltpqvmnrpsissw-----dglpgklytlvltqdpapsrkdpkfrehw 85
 Qy 112 HWLVTDIKADLUKGIQOELSAIQAPSPPAHSGFHRVQFVYLQEGKVIS----LLPK 167
 Db 86 hflvnmkgndissgtv-----lseyvggppkdtglhrvylvlyeqeqpincdepilnsk 141
 Qy 168 ENKTRGSKWMDRFLNRFHLGPEPEASTQFMTQNYQDS 203
 Db 142 sgdnrgkfkesfrkkyhlgapvagtcfqae-wdds 176

RESULT 6
 R49942
 ID R49942 standard; Protein; 187 AA.
 XX AC R49942;
 XX DT 13-OCT-1994 (first entry)
 XX DE Rat hippocampal cholinergic neurotrophic peptide precursor.
 XX KW Rat hippocampal cholinergic neurotrophic peptide; rat HCNP;
 XX KW nerve degeneration; acetylcholine synthesis; neurostimulation.
 XX OS Rattus norvegicus (Wistar).
 XX FH Key Location/Qualifiers
 XX FT Peptide 2..12
 XX FT Protein /label= rat_HCNP
 XX FT /label= precursor
 XX PN WO9405788-A.
 XX PD 17-MAR-1994.
 XX PF 27-AUG-1993; 93WO-JP01214.
 XX PR 29-AUG-1992; 92JP-0253734.
 XX PR 28-AUG-1992; 92JP-0254170.
 XX PA (SUMU) SUMITOMO PHARM CO LTD.
 XX PA (YAMA/) YAMAMOTO M.
 XX PI Agui H, Fukushima N, Irie T, Kamikawa Y, Kojima S;
 XX PI Nishihara T, Ojika K, Ono K, Taiji M, Tohdoh N;
 XX PI Tojo S, Ueki Y;
 XX DR WPI; 1994-101193/12.
 XX DR N-PSDB; Q58685.
 XX PT Neurotrophic peptide(s), precursors and genes - used to treat
 XX PT nervous degeneration, increases acetylcholine synthesis
 XX PS Claim 3; Page 155-156; 200pp; Japanese.
 XX CC The first 11 amino acids from the mature N-terminus of the rat
 XX CC hippocampal cholinergic neurotrophic peptide precursor have
 XX CC neurotrophic activity and increase acetylcholine synthesis.
 XX SQ Sequence 187 AA;

Query Match 19.2%; Score 235.5; DB 15; Length 187;
 Best Local Similarity 36.5%; Pred. No. 6.6e-19;
 Matches 57; Conservative 20; Mismatches 60; Indels 19; Gaps 5;

Qy 52 ELGNIGCKVVDPCNNYRKITSWMEPIVKFPGAVDGTIYLVWVDPDAPSRAEPRQRFWR 111
 Db 36 elgkv---ltpqvmnrpsissw-----dglpgklytlvltqdpapsrkdpkfrehw 85
 Qy 112 HWLVTDIKADLUKGIQOELSAIQAPSPPAHSGFHRVQFVYLQEGKVIS----LLPK 167

Db 61 vpdxcnyrqktsmepivkfpagvgatylvmvdpapsraepqrgrfwrhwlvtidkg 120
 Qy 121 A 121
 Db 121 a 121

RESULT 5
 R27897
 ID R27897 standard; Protein; 187 AA.
 XX AC R27897;
 XX DT 18-MAR-1993 (first entry)
 XX DE HCNP protein.
 XX KW Hippocampal cholinergic neurotrophic peptide; HCNP; hippocampal tissue;
 XX KW neonatal rat; expression vector; neurodegenerative disorder; dementia;
 XX KW Alzheimer's disease; Parkinson's disease.
 XX OS Rattus norvegicus.
 XX FH Key Location/Qualifiers
 XX FT Peptide 2..12
 XX FT /label= HCNP
 XX FT Misc-difference 135
 XX FT /note= "May opt. be Lys"
 XX PN EP511816-A.
 XX PD 04-NOV-1992.
 XX PF 27-APR-1992; 92EP-0303800.
 XX PR 27-APR-1991; 91JP-0124688.
 XX PA (OJIK/) OJIK K.
 XX PA (SUMU) SUMITOMO PHARM CO LTD.
 XX PA (YAMA/) YAMAMOTO M.
 XX PI Agui H, Fukushima N, Irie T, Kojima S, Nishihara T;
 XX PI Ojika K, Ono K, Tohdoh N, Tojo S, Ueki Y;
 XX DR WPI; 1992-367633/45.
 XX DR N-PSDB; Q30001.
 XX PT Neurotrophic peptide derivs. - used for treating neurological
 XX PT degenerative disorders, e.g. Alzheimer's disease or Parkinson's
 XX PT disease
 XX PS Claim 1 and claim 3; Page 31-32; 57pp; English.
 XX CC The sequence given is the hippocampal cholinergic neurotrophic peptide
 XX CC (HCNP). The active part of this protein is located in the first
 XX CC eleven amino acids at the N terminus. The gene encoding this sequence
 XX CC was obtained by preparing mRNA from the hippocampal tissue of neonatal
 XX CC rats, 12 days after birth. The mRNA is converted to cDNA by standard
 XX CC methods. The cDNA can be used in expression vectors to produce both
 XX CC the full length protein and the active peptide fragment. The gene and
 XX CC the peptide are useful for the treatment of neurodegenerative
 XX CC disorders and dementia eg. Alzheimer's and Parkinson's disease.
 XX SQ Sequence 187 AA;

Query Match 19.2%; Score 235.5; DB 13; Length 187;
 Best Local Similarity 36.5%; Pred. No. 6.6e-19;
 Matches 57; Conservative 20; Mismatches 60; Indels 19; Gaps 5;

Qy 52 ELGNIGCKVVDPCNNYRKITSWMEPIVKFPGAVDGTIYLVWVDPDAPSRAEPRQRFWR 111
 Db 36 elgkv---ltpqvmnrpsissw-----dglpgklytlvltqdpapsrkdpkfrehw 85


```
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
ID R49943
XX AC R49943;
XX DE 13-OCT-1994 (first entry)
XX DT Human hippocampal cholinergic neurotrophic peptide precursor.
XX DE Human hippocampal cholinergic neurotrophic peptide; human HCNP;
XX KW nerve degeneration; acetylcholine synthesis; neurostimulation.
XX KW Homo sapiens.
XX OS
XX FH Key Location/Qualifiers
XX FT Peptide 2..12
XX FT /label= human_HCNP
XX FT Protein 1..187
XX FT /label= precursor
XX PN W09405788-A.
XX PD 17-MAR-1994.
XX PF 27-AUG-1993; 93WO-JP01214.
XX PR 29-AUG-1992; 92JP-0253734.
XX PR 28-AUG-1992; 92JP-0254170.
XX PA (SUMU) SUMITOMO PHARM CO LTD.
XX PA (YAMA) YAMAMOTO M.
XX PI Agui H, Fukushima N, Irie T, Kamikawa Y, Kojima S;
XX PI Nishihara T, Ojika K, Ono K, Taiji M, Tohdoh N;
XX PI Tojo S, Ueki Y;
XX DR WPI: 1994-101193/12.
XX DR N-PSDB; Q58666.
XX PT Neurotrophic peptide(s), precursors and genes - used to treat
XX PT nervous degeneration, increases acetylcholine synthesis
XX PS Claim 4; Page 164-166; 200pp; Japanese.
XX CC The first 11 amino acids from the N-terminus of the human
XX CC hippocampal cholinergic neurotrophic peptide precursor have
XX CC neurotrophic activity and increase acetylcholine synthesis.
XX SQ Sequence 187 AA;

Query Match 16.7%; Score 205; DB 15; Length 187;
Best Local Similarity 33.1%; Pred. No. 2e-15;
Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRQKITSWMEPIVFPFGAVDGTATYILVMVDPDAPSPRAEPRQRFWR 111
DB 36 elgkv---ltptqvknpptsisw-----dglsgklytlvtldpapsrkdkpyrewh 85
QY 112 HWLVTDIKGADLKKIGQOELSAYQAPSPPAHSGFHRVQFFVYLOEGKVIS----LLPK 167
DB 86 hflvnmkgndissgtv----lsdyvsgppkggtghryvwlvyeqdrplkcdpilsnr 141
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
DB 142 sgdhrgkfkwaskfrkkyelrapvagtcygae 172
```

```
RESULT 10
R64268
ID R64268 standard; Protein; 187 AA.
XX AC R64268;
XX DT 04-JUL-1995 (first entry)
XX DE Phosphatidylethanolamine binding protein.
XX KW Human; phosphatidylethanolamine binding protein; hPEBP; cDNA library;
XX KW glioblastoma; T98G; aplasia; proliferation; glia; neuron;
XX KW hematopoietic cell; immunological activity; neurological activity;
XX KW inflammatory disease; tumour; antibody.
XX OS Homo sapiens.
XX PN EP628631-A.
XX PD 14-DEC-1994.
XX PF 30-MAY-1994; 94EP-0108321.
XX PR 08-JUN-1993; 93JP-0137042.
XX PA (ONOF) ONO PHARM CO LTD.
XX PI Naito T, Nakade S;
XX DR WPI: 1995-015698/03.
XX DR N-PSDB; Q80734.
XX PT New human phosphatidylethanolamine binding protein gene - used to
XX PT develop prods. for diagnosis and treatment of diseases involving
XX PT glia, neurons and other cells.
XX PS Claim 1; Page 13-14; 17pp; English.
XX CC This sequence represents the human phosphatidylethanolamine binding
XX CC protein, hPEBP. The cDNA was isolated from a cDNA library prepared
XX CC using RNA extracted from the human glioblastoma cell line T98G. This
XX CC protein can be used for the prevention of or in the treatment of
XX CC aplasia or abnormal proliferation of glia, neurons or hematopoietic
XX CC cells, depression or enhancement of immunological or neurological
XX CC activity, inflammatory disease, tumours or diseases induced by
XX CC abnormal lipid metabolism. Anti-PEBP antibodies may be used for
XX CC diagnosis of genetic diseases.
XX SQ Sequence 187 AA;

Query Match 16.7%; Score 205; DB 16; Length 187;
Best Local Similarity 33.1%; Pred. No. 2e-15;
Matches 50; Conservative 21; Mismatches 62; Indels 18; Gaps 4;

QY 52 ELGNIGCKVVPDCNNYRQKITSWMEPIVFPFGAVDGTATYILVMVDPDAPSPRAEPRQRFWR 111
DB 36 elgkv---ltptqvknpptsisw-----dglsgklytlvtldpapsrkdkpyrewh 85
QY 112 HWLVTDIKGADLKKIGQOELSAYQAPSPPAHSGFHRVQFFVYLOEGKVIS----LLPK 167
DB 86 hflvnmkgndissgtv----lsdyvsgppkggtghryvwlvyeqdrplkcdpilsnr 141
QY 168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMTO 198
DB 142 sgdhrgkfkwaskfrkkyelrapvagtcygae 172

RESULT 11
Y49098
ID Y49098 standard; Protein; 175 AA.
XX
```


FT Modified-site 56..58 /note= "N-linked glycosylation site"
FT Modified-site 61..63 /note= "N-linked glycosylation site"
FT Modified-site 119..121 /note= "N-linked glycosylation site"
FT Modified-site 124..126 /note= "N-linked glycosylation site"

XX US7644372-A.
XX 05-NOV-1991.
XX 23-JAN-1991; 91US-0644372.
XX 23-JAN-1991; 91US-0644372.
XX (USSH) NAT INST OF HEALTH.
XX Lazzeri MESL, Nutman TB, Weiss N;
XX WPI; 1991-361687/49.
XX N-PSDB; Q14832.
XX Onchocerca volvulus antigen (OV-16) and DNA - useful in early and
PT specific diagnosis of onchocerciasis
XX Disclosure: Fig 4; 48pp; English.

XX The amino acid sequence is that of the Onchocerca volvulus antigen
XX (OV-16) which may be used in the diagnosis of onchocerciasis in an
XX animal. Early and specific diagnosis of new or re-infections with O.
XX volvulus in vector reinvasion areas, as well as the detection of light
XX infections in areas where control is being attempted by widespread use
XX of ivermectin, is possible. This allows for monitoring, evaluating and
XX consolidating onchocerciasis control by both the vector control and
XX chemotherapeutic strategies.
XX Sequence 152 AA;

Query Match 11.4%; Score 139; DB 12; Length 152;
Best Local Similarity 38.5%; Pred. No. 4.9e-08;
Matches 35; Conservative 9; Mismatches 33; Indels 14; Gaps 4;
QY 55 NGCKVVP-DCNNYRKQKTSWMEPIVKFGVDGATYILVMVDPDAPSRAEPRQRFRWH 113
Db 65 nlgneltptqvknpktkv-sw-----daegpalytlvmtddapsrknvpvfrwhhw 115
QY 114 LVTDIKGDADLKGKIQGQELSAYOAPSPPAH 144
Db 116 llinisgqnvsgtv---lsdywirstkrh 142

RESULT 14
W13945
ID W13945 standard; Protein; 177 AA.
XX AC W13945;
XX 24-JUN-1997 (first entry)
XX Arabidopsis terminal flower 1 protein.
XX Arabidopsis terminal flower 1 protein.
XX Terminal flower 1; tfl1 gene; centroradialis protein; cen gene;
KW flowering; transgenic plant.
XX Arabidopsis thaliana var. Columbia.
XX OS W09710339-A1.
XX 20-MAR-1997.
XX

PF 13-SEP-1996; 96WO-GB02276.
XX 13-SEP-1995; 95GB-0018731.
XX (INNE-) INNES CENT JOHN.
XX Bradley DJ, Carpenter R, Coen ES;
PI WPI; 1997-202235/18.
XX N-PSDB; T60140;
XX N-PSDB; T60142.

XX Antirrhinum majus centroradialis gene and Arabidopsis homologue,
PT tfl1 - control switching of apical meristem to floral fate, useful
PT in plant hybridisation and in control of growing season
XX Claim 12; Fig 6a; 83pp; English.

XX The terminal flower 1 (TFL1) protein (W13945) of Antirrhinum majus
CC plays a role in preventing the apical meristem from switching to
CC a floral fate and in inhibition of flowering. Its sequence can be
CC deduced from an isolated cDNA clone (T60140) and genomic sequence
CC (T60142). Mutants and variant TFL1 polypeptides, and homologous
CC CSN protein (see also W13944) of Antirrhinum, can be expressed in
CC transgenic plants in order to influence flowering characteristics,
CC esp. the switching of apical meristem to a floral fate and the
CC timing of flowering. Antibodies raised against the polypeptides
CC are useful in the identification and isolation of homologous
CC proteins.
XX Sequence 177 AA;

Query Match 11.2%; Score 137.5; DB 18; Length 177;
Best Local Similarity 30.2%; Pred. No. 8.9e-08;
Matches 38; Conservative 25; Mismatches 50; Indels 13; Gaps 5;
QY 76 EPIVKFPGAVDGYIILVMVDPDAPSRAEPRQRFRWHLVTDIKG-ADLKKGKIQGQELS 134
Db 54 kprveihggdlrsfftlvmidpdpvpsdpflkehlhwtvtnipgtdatfgk----evv 109

QY 135 AYQAPSPPAHSGFHRHYOFFVYLOEGK--VISLLPKENKTRGSKMDFLNRFLHLEPEAS 192
Db 110 syelprpsi--glhrfvfvfrqkrvifpnip----srdfntkrfaveydlglpva 163
QY 193 TQFMTQ 198
Db 164 vffnaq 169

RESULT 15
Y11503
ID Y11503 standard; Protein; 120 AA.
XX AC Y11503;
XX 21-JUN-1999 (first entry)
XX Human 5' EST secreted protein SEQ ID No 325.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX Homo sapiens.
XX W09906551-A2.
XX 11-FEB-1999.
XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:25:43 ; Search time 73.46 seconds
(without alignments)
362.186 Million cell updates/sec

Title: US-09-215-435-225
Perfect score: 1224
Sequence: 1 MGWTRLVTAALLGLMMV.....APRERASEPKHKNAETAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229	18.7	178	5 Q9VK60	Q9VK60 drosophila
2	225	18.4	187	11 Q9J58	Q9J58 mus musculus
3	222.5	18.2	187	5 Q9VT08	Q9VT08 drosophila
4	216.5	17.7	202	5 Q9VI09	Q9VI09 drosophila
5	206.5	16.9	173	10 Q9XF7	Q9XF7 arabidopsis
6	196	16.0	171	5 Q9NKY4	Q9NKY4 dirofilaria
7	186.5	15.2	179	5 Q9VD02	Q9VD02 drosophila
8	183	15.0	172	5 Q9XW37	Q9XW37 caenorhabdi
9	182	14.9	175	10 Q9S7R5	Q9S7R5 arabidopsis
10	182	14.9	210	5 Q9VP29	Q9VP29 drosophila
11	179	14.6	177	10 Q9SXY9	Q9SXY9 citrus unsh
12	175.5	14.3	176	5 Q9VD01	Q9VD01 drosophila
13	164	13.4	174	10 Q9XH44	Q9XH44 nicotiana t
14	164	13.4	175	10 Q04467	Q04467 arabidopsis
15	159	13.0	182	5 Q9YIK8	Q9YIK8 octopus dof
16	154.5	12.6	173	10 Q9XGS4	Q9XGS4 oryza sativ
17	152	12.4	173	10 Q9XGS5	Q9XGS5 oryza sativ
18	148.5	12.1	175	10 Q82088	Q82088 lycopersico
19	147.5	12.1	175	10 Q9XH43	Q9XH43 nicotiana t

20	147	12.0	175	10 Q9ZNV5	Q9ZNV5 arabidopsis
21	144.5	11.8	175	10 Q9XH42	Q9XH42 nicotiana t
22	138	11.3	118	10 Q9XH41	Q9XH41 nicotiana t
23	137.5	11.2	177	10 P93003	P93003 arabidopsis
24	134.5	11.0	177	10 Q9SAY4	Q9SAY4 brassica ol
25	134.5	11.0	178	10 Q82152	Q82152 brassica na
26	134.5	11.0	178	10 Q9SAY5	Q9SAY5 brassica na
27	132.5	10.8	178	10 Q82153	Q82153 brassica ra
28	132.5	10.8	178	10 Q9XGD4	Q9XGD4 brassica ca
29	122.5	10.0	178	10 Q9SAY6	Q9SAY6 brassica na
30	115.5	9.4	177	10 Q82154	Q82154 brassica ol
31	113	9.2	215	3 Q9VP69	Q9VP69 neurospora
32	103	8.4	409	5 Q9V748	Q9V748 drosophila
33	101	8.3	122	10 Q9SXZ2	Q9SXZ2 arabidopsis
34	96.5	7.9	150	2 Q9JO64	Q9JO64 chlamydia p
35	93.5	7.6	157	1 Q9UXF4	Q9UXF4 sulfolobus
36	93.5	7.6	186	1 Q9UJ33	Q9UJ33 pyrococcus
37	93	7.6	180	2 Q9KT30	Q9KT30 vibrio chol
38	90.5	7.4	209	2 Q9KJE6	Q9KJE6 thauera aro
39	89.5	7.3	1945	5 Q96681	Q96681 drosophila
40	89.5	7.3	1945	5 Q9V491	Q9V491 drosophila
41	88	7.2	174	4 Q9P0B9	Q9P0B9 homo sapien
42	87	7.1	150	2 Q9PLJ0	Q9PLJ0 chlamydia m
43	85	6.9	449	4 Q9UJH6	Q9UJH6 homo sapien
44	85	6.9	595	4 Q9Y5X1	Q9Y5X1 homo sapien
45	85	6.9	2253	13 P70012	P70012 xenopus lae

ALIGNMENTS

RESULT	1				
Q9VK60					
ID	Q9VK60	PRELIMINARY;	PRT;	178 AA.	
AC	Q9VK60;				
DT	01-MAY-2000 (TReMBLrel. 13, Created)				
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)				
DE	01-OCT-2000 (TReMBLrel. 15, Last annotation update)				
DE	CG6180 PROTEIN.				
GN	CG6180.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BERKELEY;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Ananulides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,				
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				

RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M.	Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of Drosophila melanogaster.";
RT	Science 287:2185-2195(2000).	
RL	EMBL; AE003637; AAF5321.1; -.	
DR	HSSP; PI3696; IA44.	
DR	FLYBASE; FBgn0032453; CG6180.	
DR	INTERPRO: IPR001858; -.	
DR	PFAM; PF01161; PBP; -.	
SQ	SEQUENCE 178 AA; 19707 MW; EEB29266CFEDB1E6 CRC64;	
	Query Match 18.7%; Score 229; DB 5; Length 178;	
	Best Local Similarity 40.5%; Pred. No. 3.2e-15;	
	Matches 53; Conservative 17; Mismatches 51; Indels 10; Gaps	
QY	76 EPIVKPGAVDGYTILVWDDPADSPRSPRQRFWRHVLVTDIKGADLKKGIQGELSA	135
DB	: : :	
DB	43 EPCVKWE-ADANKLYTLCMTDPDPSRKDKPKFEWHHLVGVNIPGDVAKEV----	LSA 97
QY	136 YQAPSPPAHSFGHYRYFFVYLQGGKYI---SLLPKEN-KTRGSWKMDRFLNRFH	191
DB	: : :	
DB	98 YVSGSPPDTGLHRYVFLIVEQRCKLTDFDEKRLPNNSGDSGRGFKTAIEAKK	VALGNPIA 157
QY	192 STQFMNTQNYQD 202	
DB	:	
DB	158 GNLYQAE-YDD 167	
RESULT 2		
ID	Q9JJ58 PRELIMINARY; PRT; 187 AA.	
AC	Q9JJ58;	
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)	
DE	HIPPOCAMPAL CHOLINERGIC NEUROSTIMULATING PEPTIDE PRECURSOR PROTEIN.	
GN	HCNPPP.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
XC	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Matsukawa N.;	
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
[2]		
RP	SEQUENCE FROM N.A.	
RA	Matsukawa N., Tooyama I., Kimura H., Yamamoto T., Yasukuni T., Oomura Y., Ojika K.;	
RT	"Increased expression of hippocampal cholinergic neurostimulating peptide-related components and their messenger RNAs in the hippocampus of aged senescence-accelerated mice."	
RL	Neuroscience 88:79-92(1999).	
DR	EMBL; AB046417; BAB03276.1; -.	
SQ	SEQUENCE 187 AA; 20830 MW; 4059C7C6E4A8BAA0 CRC64;	
	Query Match 18.4%; Score 225; DB 11; Length 187;	
	Best Local Similarity 35.1%; Pred. No. 8.7e-15;	
	Matches 53; Conservative 22; Mismatches 58; Indels 18; Gaps	
OY	52 ELGNIGCKVVPDCNNYRKQITSWNEPIVKFPFGVDGATYILVMVDPDAPSRAEPQBFR	111

DR PROSITE; PS01220; PBP; 1.
SQ SEQUENCE 187 AA; 20872 MW; A3CB65A04EEACC01 CRC64;

Query Match 18.2%; Score 222.5; DB 5; Length 187;
Best Local Similarity 38.6%; Pred. No. 1.5e-14;
Matches 49; Conservative 21; Mismatches 48; Indels 9; Gaps 3;

QY 76 EPIVKKPGAVGATYILVMVDPPAPRAEPRQRFWRHLVTDIKGADLKKKGIOGELSA 135
DB 50 QPKVKW-DADPNFYLLTLLDPPAPRRKKPKFKEWHHLVNPQGVGVV-----LTE 104

QY 136 YQAPSPPAHSGFHYQFFVYLOEGKVISLPPKENTK-----RGSWKMDRFLNRHFLGEPEA 191
DB 105 YVGAGPQGGLHRYVFLFKQPKLTKNEPKIPKTSKDRANFSTSKFNSKYLKLGDP 164

QY 192 STQFMQ 198
DB 165 GNFFQA 171

RESULT 4
Q9VI09 PRELIMINARY; PRT; 202 AA.
AC Q9VI09;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CGL7919 PROTEIN.
GN CGL7919.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flesler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ithegawam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Minkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AB003675; AAF54136.1; -.
DR HSSP; P30086; 1BEH.
DR FLYBASE; FBgn0037433; CG17919.
DR INTERPRO; IPR001858; -.
DR PFAM; PF01161; PBP; 1.
DR PROSITE; PS01220; PBP; 1.
SQ SEQUENCE 202 AA; 22642 MW; F0333A13597AC596 CRC64;

Query Match 17.7%; Score 216.5; DB 5; Length 202;
Best Local Similarity 28.0%; Pred. No. 6.8e-14;
Matches 59; Conservative 31; Mismatches 72; Indels 49; Gaps 6;

QY 13 LLGLMMVVTGDEIDENSPCAHEALDEDTLFCQGLEVFYFPELGNIGCKVVPDCNNYRKIT 72
DB 8 LVGCLLAVQAGSVVEVFRSHQVVD-----VIPEPPNQLKVT 45

QY 73 SWM-----EPIVKFPGAVDGYIILVMVDPPAPRAEPRQRFWRHLV 115
DB 46 YSNNLVAKDGVLTPTQVKDQPVVW-DAQPGFEYTLTMDPPAPRAEPRQRFWRHLV 104

QY 116 TDIKGADLKKKGIOGELSAVQAPSPPAHSGFHYQFFVYLOEGKV-----ISLLPKENKT 171
DB 105 ANIAGNDL-----ASGEPIAEYIGSGPQGTGLHRYVFLLYKQSGKLEFDEERSKRSKRD 160

QY 172 RGSWKMDRFLNRHFLGEPEASTQFMQYQD 202
DB 161 RPKFSAKFAINHELGNPIAGTFYQAQ-YDD 190

RESULT 5
Q9XFK7 PRELIMINARY; PRT; 173 AA.
AC Q9XFK7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE EL2A11 PROTEIN (T10F20.11 PROTEIN) (T10022.8).
GN EL2A11 OR T10F20.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Sakano H., Liu S.X., Etgu P., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA Yu G., Chan A., Chung M., Goldsmith A., Liu A., Smith A., Vaysberg M.,
RA Altatfi H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B.,
RA Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,
RA Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Federspiel N.A., Theologis A.;
RT "The sequence of BAC T10F20 from Arabidopsis thaliana chromosome 1.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altatfi H., Bel Q., Chin C., Chiu J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu S., Mukharsy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC T10022 from chromosome I."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF147721; AAD37380.1; -
 DR EMBL: AC034107; AAF97827.1; -
 DR EMBL: AC069551; AAF78390.1; -
 DR HSSP: Q41261; I000.
 DR INTERPRO: IPR001858; -
 DR PFAM: PF011161; PBP; 1.
 DR PROSITE: PS01220; PBP; 1.
 SQ SEQUENCE 173 AA; 19134 MW; 3D548E21AD959A54 CRC64;

Query Match 16.9%; Score 206.5; DB 10; Length 173;
 Best Local Similarity 34.1%; Pred. No. 5.6e-13;
 Matches 56; Conservative 19; Mismatches 68; Indels 21; Gaps 7;
 QY 49 FYPELGNICKVVPDCNNYRQKITSMETIVKFPAGVAGCATYILVWVDPDAPSRAEPROR 108
 DB 31 FGPKHNTNGCEIKPS-----TAVNPPKVNISGHSDELTLVWVDPDAPSSEPNNR 81
 QY 109 FWRHRLVLTDKGADLKKGKTKGQELSAYQAPSPPAHSGFHRFYQFFVYLOEGKVISLLPK 168
 DB 82 EWHVNIWVDIPGG--TNSRGKEILPYMEPPV--GIHRY-ILVLFRRQSPVGLMWQO 135
 QY 169 NKTGRGSKMDFLNRHFLGPEASTQFMNTQYDSTLQAPRER 212
 DB 136 PPSRANFSTRMFAGHFDGLGLPVATVYF---NAQKEP---ASRRR 173

RESULT 6
 ID Q9NKY4 PRELIMINARY; PRT; 171 AA.
 AC Q9NKY4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE PHOSPHATIDYL-ETHANOLAMINE-BINDING PROTEIN.
 OS Dirofilaria immitis.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Dirofilaria.
 OC NCBI_TaxID=6287;
 RN
 RP SEQUENCE FROM N.A.
 RA Kawazu S., Sato K., Sasaki Y., Nishiyama A., Noguchi M., Tsuji N.,
 RA Suzuki M., Kano S.;
 RA "Dirofilaria immitis mRNA for phosphatidyl-ethanolamine-binding
 RT protein, complete cds."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB035321; BAA96354.1; -
 DR EMBL: AB035321; BAA96354.1; -
 SQ SEQUENCE 171 AA; 19293 MW; FLE2793A30D4254A CRC64;

Query Match 16.0%; Score 196; DB 5; Length 171;
 Best Local Similarity 34.8%; Pred. No. 6.2e-12;
 Matches 39; Conservative 26; Mismatches 45; Indels 2; Gaps 2;

QY 88 ATYILVWVDPDAPSRAEPRORFWRHRLVLTDKGADLKKGKTKGQELSAYQAPSPPAHSGF 147
 DB 60 STFSMIMIDPNLRSKNPSVAEWLHVLVNPISNIEGGHGOHQWAGSPAPQPRDI 119
 QY 148 HRYQFFVYLOEGKVISLLPKENKTRGSKMDFLNRHFLGPEASTQFMNTQ 199
 DB 120 HRYIILLYEHGRRIQ-VPKIN-SRAKNFIQKVEKHGLGDPITAGNFLEAQN 169

RESULT 7
 ID Q9VD02 PRELIMINARY; PRT; 179 AA.
 AC Q9VD02;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE CG7054 PROTEIN.
 GN CG7054.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RP
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Pan S.,
 RA Palazzolo M., Pittman G.S., Pan S., Poirard J., Puri V., Reese M.G.,
 RA Relnert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003739; AAF56003.1; -
 DR HSSP: P13696; 1A44.
 DR FLYBASE: FBgn0038972; CG7054.
 DR INTERPRO: IPR001858; -
 DR PFAM: PF01161; PBP; 1.
 SQ SEQUENCE 179 AA; 19822 MW; DIC2046AEBFA6603 CRC64;

Query Match 15.2%; Score 186.5; DB 5; Length 179;
 Best Local Similarity 35.6%; Pred. No. 5.8e-11;
 Matches 47; Conservative 23; Mismatches 53; Indels 9; Gaps 5;
 QY 76 EPVIVKPGAVDGTATYI--LVWVDPDAPSRAEPRORFWRHRLVLTDKGADLKKGKTKGQEL 133
 DB 41 OPTVWSWG-LEGKSNLJLTLLMVDPDAPTQDPKYRIILHWSVNPVPGSN--ENPSGGHSL 97
 QY 134 SAYQAPSPPAHSGFHRFYQFFVYLOEGKVISLLPKENKTRG---SWKMDFLNRHFLGPE 190
 DB 98 ADYVGSQPKDTGLHRYIFLLYRQENKIETPTISNTRTGRNLNFARNDAFAKHGLGEPT 157
 QY 191 ASTQFMNTQYD 202
 DB 158 AANYQAO-YDD 168

Db 65 YTLVMVDPVSPSPNPHLEHLVLTDPAT---TGTFGNEIVCYENPSPTA--GIHR 119
QY 150 YQFFVYLQGGKVISLLPKENKTRGSKMD----RFLNRFHLGPEASTQFMTO 198
Db 120 VVFLPRQLGRQTVVAP-----GWRQNFRTFAEINLGLPVAAVFYNCQ 165

RESULT 15
Q9YIK8 PRELIMINARY; PRT; 182 AA.
AC Q9YIK8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE O-CRYSTALLIN.
OS Octopus dofleini (Giant octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Colecoidea; Octopoda;
OC Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=6644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99280325; PubMed=10350626;
RA Zinovieva R.D., Platiogorsky J., Tomarev S.I.;
RT "O-Crystallin, arginine kinase and ferritin from the octopus lens."
RL Biochim. Biophys. Acta 1431:512-517(1999).
DR EMBL; AF117272; RAD29640.1; -;
DR HSSP; Q41261; IQOU.
DR INTERPRO; IPR001858; -;
DR PFAM; PF01161; PBP; 1.
SQ SEQUENCE 182 AA; 21241 MW; D75DB3CC087B22AD CRC64;

Query Match 13.0%; Score 159; DB 5; Length 182;
Best Local Similarity 30.7%; Pred. No. 3.3e-08;
Matches 43; Conservative 25; Mismatches 48; Indels 24; Gaps 6;
QY 76 EPIVKEPFGAVDGYTLVMVDPDAPSRAPRQRFWRHRLVTDIKGADLKKKIQGOELSA 135
Db 45 QPQIKFE-AETNYYIILNDADFPSRSQKLNFEQHLVWVNIPIGSDISRGDV----LTD 99
QY 136 YQAPSPPAHSGFHYQFFVYLQ-----EGKVISLLPKENKT---RGSWKMDRFLNR 183
Db 100 YIGPLPNKGTGYHRYVLMFLFKQSKGRMEFRGEKKI-----NNRTSEGRKSYNMMFEFARK 153
QY 184 FHLGPEASTQFMTOYQDS 203
Db 154 HFLVPEYIGNFFQSE-WDDS 172

Search completed: March 24, 2001, 13:25:45
Job time: 251 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:16 ; Search time 44.26 Seconds
(without alignments)
348.248 Million cell updates/sec

Title: US-09-215-435-225

Perfect score: 1224

Sequence: 1 MGWTMLVTAALLGLLMVV.....APRRASEPKHKNAQIAAC 227

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235.5	19.2	187	2 A36126	23K phosphatidylylet
2	234.5	19.2	221	2 T31721	hypothetical prote
3	223	18.2	134	2 PC4214	phosphatidylylethano
4	222	18.1	152	2 PC4216	phosphatidylylethano
5	218.5	17.9	186	2 S00056	basic cytosolic pr
6	218.5	17.9	186	2 S18358	23K protein - rat
7	210	17.2	114	2 PC4215	phosphatidylylethano
8	206	16.8	187	2 S46485	phosphatidylylethano
9	205	16.7	187	2 T53745	phosphatidylylethano
10	183	15.0	172	2 T27310	hypothetical prote
11	182	14.9	175	2 T10579	TfL1 protein homol
12	169.5	13.8	262	2 A57391	TcSL-2 protein pre
13	148.5	12.1	175	2 T07877	self-pruning prote
14	148	12.1	181	2 S68507	centroradialis pro
15	139	11.4	152	2 A44947	16k antigen precu
16	137.5	11.2	177	2 T48411	terminal flower1 (
17	133	10.9	201	2 S51424	hypothetical prote
18	122.5	10.0	219	2 S18843	TfS1 protein - yea
19	113.5	9.3	197	2 C70519	hypothetical prote
20	113	9.2	151	2 E69134	conserved hypothet
21	113	9.2	215	2 T48789	TfS1 related prote
22	106	8.7	171	2 B70408	conserved hypothet
23	105	8.6	198	2 B71072	hypothetical prote
24	96.5	7.9	150	2 C72024	conserved hypothet
25	95	7.8	183	2 G64786	ybCL protein - Esc
26	95	7.8	201	2 D70519	probable lppc prot
27	93.5	7.6	186	2 C75095	hypothetical prote
28	93	7.6	180	2 C82243	conserved hypothet
29	92	7.5	179	2 T36855	conserved hypothet

ALIGNMENTS

RESULT 1

A36126

23K phosphatidylylethanolamine-binding protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Mar-1991 #sequence-revision 28-Mar-1991 #text_change 20-Jun-2000

C:Accession: A36126; S46484; S37556; I84473; S37555

R:Grandy, D.K.; Hanneman, E.; Bunzow, J.; Shih, M.; Machida, C.A.; Bidlack, J.M.; Cliv Mol. Endocrinol. 4, 1370-1376, 1990

A:Title: Purification, cloning, and tissue distribution of a 23-kDa rat protein isola

A:Reference number: A36126; MUID:91042640

A:Accession: A36126

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-187 <GRA>

R:Perry, A.C.F.; Hall, L.; Bell, A.E.; Jones, R.

Biochem. J. 301, 235-242, 1994

A:Title: Sequence analysis of a mammalian phospholipid-binding protein from testis an

A:Reference number: S46484; MUID:94311839

A:Accession: S46484

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-187 <PER>

A:Cross-references: EMBL:X71873; NID:g510338; PIDN:CAA50708.1; PID:g510339

R:Tohdo, N.; Agui, H.; Ojika, K.

submitted to the EMBL Data Library, September 1993

A:Description: Sequences of rat HCNP precursor protein promoter region.

A:Reference number: S37556

A:Accession: S37556

A:Molecule type: DNA

A:Residues: 1-45 <TOD>

A:Cross-references: EMBL:X75254

R:Tohdo, N.; Tojo, S.; Agui, H.; Ojika, K.

Brain Res. Mol. Brain Res. 30, 381-384, 1995

A:Title: Sequence homology of rat and human HCNP precursor proteins, bovine phosphati

A:Reference number: I38109; MUID:95364631

A:Accession: I84473

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-187 <RES>

A:Cross-references: EMBL:X75253; NID:g406291; PIDN:CAA53032.1; PID:g406292

C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 19.2%; Score 235.5; DB 2; Length 187;
Best Local Similarity 36.5%; Pred. No. 3.4e-16;
Matches 57; Conservative 20; Mismatches 60; Indels 19; Gaps 5;

Oy 52 ELGNIGCKVDPCCNNYRQKITSNMEPIVKFPGAVDGYTILVMVDDPAPSRAEPRQRFWR 111

Db 36 ELGKV---LTPTQVMNRPSISW-----DGLDPGKLYTLVLTDPDAPSRRKDKFREW 85

Oy 112 HMLVTIDIKGADLKGKIQGQELSAQAPSPPAHSGFHRYQFFVYLQEGKVLS-----LLPK 167

```

Db      86 HFLVNMKNDISSGTV-----LSEYVSGPPKDTGLHRYVWLVIYEQEQLNCDEPILSNK 141
      168 ENKTRGSKMDRFLNRFHLGEPPEASTQFMQYQDS 203
      142 SGDNRGKFKVESFRKKYHLGAPVAGTCFAE-WDDS 176

RESULT  2
T31721
hypothetical protein F40A3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
C:Accession: T31721
R:Ertmann, K.D.; Bradshaw, H.; Keppler, D.
A:Description: The sequence of C. elegans cosmid F40A3.
A:Reference number: Z21074
A:Accession: T31721
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-221 <GEI>
A:Cross-references: EMBL:AF016423; PIDN:AAB65322.1; GSPDB:GN00023; CESP:F40A3.3
A:Experimental source: strain Bristol N2; clone F40A3
C:Genetics:
A:Gene: CESP:F40A3.3
A:Map position: 5
A:Introns: 32/2; 70/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match      19.2%; Score 234.5; DB 2; Length 221;
Best Local Similarity 33.1%; Pred. No. 5.3e-16;
Matches 58; Conservative 16; Mismatches 60; Indels 41; Gaps 6;

QY      59 KVVDP--CNNYRQKITSWMEPIVKFPGAVD-----GATYIL 92
      46 EVIPDLVASPPSKVVS-----VKFNSGEANLGNLTPTQKDTPEVKWDAEPGALYTL 100
QY      93 VMVDPDAPSAEPQRFWRHLVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOF 152
      101 IKYDPPDAPSRKREPTYREWHHLVVNIPGNDIAK----GDTLSEYIAGGPPPKTGLHRYVY 156
QY      153 FVYLQEGKVIS-----LLPKENKTRGSKMDRFLNRFHLGEPPEASTQFMQYQD 202
      157 LIYKQSGRIEDAEHRLTNTSGDKRGWKAADFVAKHKLGAPEVGNLFAE-YDD 210

RESULT  3
PC4214
phosphatidylethanolamine binding protein Ovd3 - nematode (Onchocerca volvulus) (fragment)
C:Species: Onchocerca volvulus
C:Date: 17-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Jun-2000
C:Accession: PC4214
R:Ertmann, K.D.; Gallin, M.Y.
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl-
Gene 174, 203-207, 1996
A:Reference number: PC4214; MUID:97045813
A:Accession: PC4214
A:Molecule type: mRNA
A:Residues: 1-134 <ERT>
A:Cross-references: EMBL:X87989; NID:g1143530; PID:g1143531
C:Genetics:
A:Introns: 34/3; 52/2; 81/1; 129/3
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match      18.2%; Score 223; DB 2; Length 134;
Best Local Similarity 36.3%; Pred. No. 4e-15;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY      55 NIGCKVVP-DCNNYRQKITSWMEPIVKFPGAVDCAATYLVVMDPDAPSRAEPQRFWRHW 113

```

```

Db      2 NLGNELTPTQVKNQTKV-SW-----DAEPGALYTLVMTDPAAPSRKPNPVRWHHW 52
QY      114 LVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOFFVYVYLGEGKVISLLPKENKTRG 173
      53 LIINISGQNVSSGTV----LSDYIGSGPKGTGLHRYVFLVYKQPGSITD--TQHGNNRP 106
QY      174 SWKMDRFLNRFHLGEPPEASTQFMQYQD 199
      107 NFKVMDFAFKHHLGNPVAAGNFFQAKH 132

RESULT  4
PC4216
phosphatidylethanolamine binding protein Ovd1 - nematode (Onchocerca volvulus) (fragment)
C:Species: Onchocerca volvulus
C:Date: 17-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Jun-2000
C:Accession: PC4216
R:Ertmann, K.D.; Gallin, M.Y.
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl-
Gene 174, 203-207, 1996
A:Reference number: PC4214; MUID:97045813
A:Accession: PC4216
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-152 <ERT>
A:Cross-references: EMBL:X87991
C:Comment: This protein is involved in odorant binding and is located in the chemosen-
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match      18.1%; Score 222; DB 2; Length 152;
Best Local Similarity 36.3%; Pred. No. 5.9e-15;
Matches 53; Conservative 18; Mismatches 59; Indels 16; Gaps 5;

QY      55 NIGCKVVP-DCNNYRQKITSWMEPIVKFPGAVDCAATYLVVMDPDAPSRAEPQRFWRHW 113
      20 NLGNELTPTQVKNQTKV-SW-----DAEPGALYTLVMTDPAAPSRKPNPVRWHHW 70
QY      114 LVTDIKGADLKKGIQGOELSAQAPSPAHSGFHRHYOFFVYVYLGEGKVISLLPKENKTRG 173
      71 LIINISGQNVSSGTV----LSDYIGSGPKGTGLHRYVFLVYKQPGSITD--TQHGNNRP 124
QY      174 SWKMDRFLNRFHLGEPPEASTQFMQYQD 199
      125 NFKVMDFAFKHHLGNPVAAGNFFQAKH 150

RESULT  5
S00056
basic cytosolic protein, 21K - bovine
N:Alternate names: phosphatidylethanolamine-binding protein
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 20-Jun-2000
C:Accession: S00056; S51191
R:Schoentgen, F.; Saccoccio, F.; Jolles, J.; Bernier, I.; Jolles, P.
Eur. J. Biochem. 166, 333-336, 1987
A:Title: Complete amino acid sequence of a basic 21-kDa protein from bovine brain cyt
A:Reference number: S00056; MUID:87275914
A:Accession: S00056
A:Molecule type: protein
A:Residues: 1-186 <SCH>
R:Bucquoy, S.; Jolles, P.; Schoentgen, F.
Eur. J. Biochem. 225, 1203-1210, 1994
A:Title: Relationships between molecular
A:Reference number: S51191; MUID:95045550
A:Contents: annotation
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5
C:Keywords: brain; cytosol; lipid binding; nucleotide binding; phospholipid
F:112-125/Region: nucleotide binding
F:168-182/Region: lipid binding

```


Query Match 17.9%; Score 218.5; DB 2; Length 186;
Best Local Similarity 33.3%; Pred. No. 1.7e-14;
Matches 53; Conservative 24; Mismatches 65; Indels 17; Gaps 4;

QY 46 LEVPI--PELGNICKVVPCDNNYRQKITSMWEPLVKFPFGAVDGATYILVMVDPPAPSRK 103
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 24 LQVKGAEVDELCKVLTPQTOKNRPSITW-----DGLDPGKLTYTLVLTDDPAPSRK 76

QY 104 EPRORFWRHMLVTDIKGADLKKGKIQQELSAYOAPSPPAHSGFHRYOFFFYVLOEGKVIS 163
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 77 DPKYREWHHLFLVNMGKNISSGV---LSDYVGSGPPKGTGLHRYVWLVEEGBPLKC 132

QY 164 ----LLPENKTROSKWMDRFNLFHLPGEPASTQFMTO 198
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 133 DEPTLSNRSGDHGRKFVASPRKKYELGAPVAGTCYQAE 171

RESULT 6
S18358
23K protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S18358
R:Jones, R.; Hall, L.
Biochim. Biophys. Acta 1080, 78-82, 1991
A:Title: A 23 kDa protein from rat sperm plasma membranes shows sequence similarity and
A:Reference number: S18358; MUID: 92031654
A:Accession: S18358
A:Molecule type: protein
A:Residues: 1-186 <JON>
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69EA.5

Query Match 17.9%; Score 218.5; DB 2; Length 186;
Best Local Similarity 33.3%; Pred. No. 1.7e-14;
Matches 53; Conservative 24; Mismatches 65; Indels 17; Gaps 4;

QY 46 LEVPI--PELGNICKVVPCDNNYRQKITSMWEPLVKFPFGAVDGATYILVMVDPPAPSRK 103
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 24 LQVKGAEVDELCKVLTPQTOKNRPSITW-----DGLDPGKLTYTLVLTDDPAPSRK 76

QY 104 EPRORFWRHMLVTDIKGADLKKGKIQQELSAYOAPSPPAHSGFHRYOFFFYVLOEGKVIS 163
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 77 DPKYREWHHLFLVNMGKNISSGV---LSDYVGSGPPKGTGLHRYVWLVEEGBPLKC 132

QY 164 ----LLPENKTROSKWMDRFNLFHLPGEPASTQFMTO 198
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 133 DEPTLSNRSGDHGRKFVASPRKKYELGAPVAGTCYQAE 171

RESULT 7
PC4215
Phosphatidylethanolamine binding protein Ovd2 - nematode (Onchocerca volvulus) (fragmented)
C:Species: Onchocerca volvulus
C:Date: 17-Dec-1996 #sequence_revision 21-Jan-1997 #text_change 20-Jun-2000
C:Accession: PC4215
R:Erttmann, K.D.; Gallin, M.Y.
Gene 174, 203-207, 1996
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidylinositol-binding protein
A:Reference number: PC4214; MUID: 97045813
A:Accession: PC4215
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-114 <ERT>
A:Cross-references: EMBL:X87990
C:Comment: This protein is involved in odorant binding and is located in the chemosensory neurons of the nematode
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69EA.5

Query Match 17.2%; Score 210; DB 2; Length 114;
Best Local Similarity 38.8%; Pred. No. 6.5e-14;
Matches 45; Conservative 14; Mismatches 51; Indels 6; Gaps 2;

RESULT 11
T10579
TFII protein homolog F9F13.20 - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 16-Jul-1999 #sequence revision 16-Jul-1999 #

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:53 ; Search time 34.59 seconds
(without alignments)
117.845 Million cell updates/sec

Title: US-09-215-435-225
Perfect score: 1224
Sequence: 1 MGWTRLVTAALLGLMMV.....APRRASEPKHKNOAETAA 227

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA: *
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2.6/ptodata/2/1aa/5_COMB.pep: *
4: /cgn2.6/ptodata/2/1aa/PctUS_COMB.pep: *
5: /cgn2.6/ptodata/2/1aa/Packfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1213	99.1	227	3	US-09-208-718-3
2	231	18.9	152	3	US-09-208-718-6
3	210.5	17.2	187	1	US-08-403-378B-4
4	205	16.7	187	1	US-08-403-378B-15
5	139	11.4	152	1	US-07-644-372-2
6	79.5	6.5	232	1	US-08-704-744-80
7	77	6.3	875	1	US-08-785-241-5
8	76	6.2	233	2	US-07-934-373C-25
9	76	6.2	233	3	US-08-437-642B-25
10	76	6.2	233	4	PCT-US93-07832-25
11	73.5	6.0	212	2	US-08-916-770-3
12	73.5	6.0	212	2	US-09-102-593-3
13	73	6.0	1049	2	US-08-817-090B-2
14	73	6.0	1050	2	US-08-817-090B-4
15	73	6.0	1101	2	US-08-916-917-14
16	73	6.0	1101	3	US-09-225-170-14
17	73	6.0	1711	2	US-08-342-930-2
18	71.5	5.8	421	2	US-08-576-626A-53
19	71.5	5.8	775	1	US-07-966-278-1
20	71.5	5.8	775	1	US-08-424-921-1
21	71.5	5.8	775	1	US-08-556-355A-1
22	71.5	5.8	775	2	US-07-803-627A-1
23	71.5	5.8	776	2	US-08-688-649-37
24	71.5	5.8	1114	2	US-08-576-626A-31
25	71	5.8	214	1	US-08-458-516-12
26	71	5.8	214	2	US-07-934-373C-39
27	71	5.8	214	3	US-08-437-642B-39
28	71	5.8	214	4	PCT-US93-07832-39

29	70	5.7	212	1	US-08-462-965A-2	Sequence 2, Appl
30	70	5.7	212	1	US-08-462-169B-21	Sequence 21, Appl
31	70	5.7	212	3	US-09-103-079-21	Sequence 21, Appl
32	70	5.7	214	1	US-08-425-763-1	Sequence 1, Appl
33	70	5.7	214	3	US-08-811-757-1	Sequence 1, Appl
34	70	5.7	1102	2	US-08-916-917-4	Sequence 4, Appl
35	70	5.7	1102	2	US-08-972-631-4	Sequence 4, Appl
36	70	5.7	1102	2	US-08-972-629-4	Sequence 4, Appl
37	70	5.7	1102	2	US-08-972-630-4	Sequence 4, Appl
38	70	5.7	1102	2	US-08-672-211-4	Sequence 4, Appl
39	70	5.7	1102	3	US-09-225-170-4	Sequence 4, Appl
40	69	5.6	214	2	US-07-934-373C-24	Sequence 24, Appl
41	69	5.6	214	3	US-08-437-642B-24	Sequence 24, Appl
42	69	5.6	214	4	PCT-US93-07832-24	Sequence 24, Appl
43	68.5	5.6	472	1	US-08-203-905B-14	Sequence 14, Appl
44	68.5	5.6	575	3	US-08-922-865-2	Sequence 2, Appl
45	68	5.6	474	2	US-08-282-197C-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1
US-09-208-718-3
; Sequence 3, Application US/09208718
; Patent No. 6063767
; GENERAL INFORMATION:
; APPLICANT: Lai, Preeti
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,718
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/958,820
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0379 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGUT12
; CLONE: 3126479
US-09-208-718-3

Query Match 99.1%; Score 1213; DB 3; Length 227;
Best Local Similarity 99.1%; Pred. No. 5.7e-131;

Matches 45; **Conservative** 34; **Mismatches** 84; **Indels** 58; **Gaps** 10;

[illegible]

RESULT 8

```

US-07-934-373C-25
; Sequence 25, Application US/07934373C
; Patent No. 5821337
; GENERAL INFORMATION:
; APPLICANT: Paul J. Carter
; APPLICANT: Leonard G. Presta
; TITLE OF INVENTION: Immunoglobulin Variants
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```

```
, ZIP: 94060  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/934, 373C  
FILING DATE: 21-Aug-1992  
CLASSIFICATION: 530  
  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/05126  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/715272  
FILING DATE: 14-JUN-1991  
  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P0709P2  
  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 233 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear
```

US-07-934-373C-25

Query Match
6.2%; Score 76; DB 2; Length 233;

Best Local Similarity 23.5%; Pred. No. 0.79;
Matches 61; Conservative 30; Mismatches 89; Indels 80; Gaps 17;

QY 1 MGWTMLVTAALLGLMMVVGTDEDE----NSPCAHEALLDDEDTLFCQGLEVFYPPELIGNI 56
||| : : | : || : || : | : | : |
Dd 1 MGW-----SCIILFLVATATGVHSDIONTPSPSLASVGDRV-----TI 40

Qy	57	GCKVPCDNN--YRQK-----ITSWMEPIV--KFGAVDGAAYILVMWDDPAP	100
Db	41	TCRASQDINNLYNQOKPKAPKPLIYYTSTLHSGVPSRFGSGSGTDYTLTI-----	94
Qy	101	SRAEPRQFRWHVLVTDIKADLKKGIQOQELS--NQAPS---PRAHSGFHYQFF	153
Db	95	SSLQOP-EDFATYYCQ---QGNLTPTTFGQGTKEIKRTVAAPSVFIFPPSDE-----	142
Qy	154	VYLQEG--KVISLL-----PKNELTRGSKMDRFLNRFHLCPEASTQFMFTQNYQDSPTLQ	207
Db	143	QLKSGTASVCLLNPNFYPREAKVQ--WKVDNALQSGNSQOEYTEQDSKDYSLSTLT	199
Qy	208	APRERASEPKHKNOAEIAAC	227
Db	200	L--SKADYEKKH-----VYAC	213

RESULT

US-08-437-042B-25
 Sequence 25, Application US/08437642B
 Patent No. 6054297
 GENERAL INFORMATION:
 APPLICANT: Paul J. Carter
 APPLICANT: Leonard G. Presta
 TITLE OF INVENTION: Immunoglobulin Variants
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 1 DNA Way
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,642B
FILING DATE: 09-May-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 07/934373
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/146206
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05126
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/715272
FILING DATE: 14-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P0709P2C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1994
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-437-642B-25

Query Match 6.2%; Score 76; DB 3; Length 233;
Best Local Similarity 23.5%; Pred. No. 0.79;
Matches 61; Conservative 30; Mismatches 89; Indels

1	QY	1	MGWTRMLVTAALLGLLMVVTGDEDE---	NSPCAHEALLDEDTLFCOGLEVFP	56
		111	: : : : : 11 :	: : : : : :	
1	dbb	1	MGW-----SCIIILFVATATGVHSDIOMTOS	SSLSASVDRV-----	40
		111	: : : : : 11 :	: : : : : :	
57	QY	57	GCKVVPDCNN-----YROK-----	ITSNMPELV--KPGGANVGATVILVMVDDP	100
		111	: : : : : 11 :	: : : : : :	
41	dbb	41	TCRASQOINVLNWOQKPKAPKLLIYVTTLSH	GVSPSRFSGSGGTDYTLT-----	94
		111	: : : : : 11 :	: : : : : :	
101	QY	101	SRAPRORFWRHVLVDIKGADLKGKIQGGELS-	-AYQAPS-----PPAHSGFHYQFP	153
		111	: : : : : 11 :	: : : : : :	
95	dbb	95	SSLQP-EDFATYYCO--OGNTLPTFGGKTV	EIKRTVAAPSVFIPPSDE-----	142
		111	: : : : : 11 :	: : : : : :	
154	QY	154	VYLOEG--KVISLL-----PKNKTRGSKMK	DFLNRHJLGEPEASTQFMQYQDSPTLQ	207
		111	: : : : : 11 :	: : : : : :	
143	dbb	143	-OLKSGTASVVCVLLNNFYPREAKVQ--	WKVDNALQSGNSQSVTEQDSDKSTYLSLT	199
		111	: : : : : 11 :	: : : : : :	
208	QY	208	APRERASEPKHNOAETAAC	227	
		111	: : : : : 11 :	: : : : : :	
200	dbb	200	L--SKADYERKH-----VYAC	213	
		111	: : : : : 11 :	: : : : : :	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 6.2% Score 76; DB 4; Length 233;
Best Local Similarity 23.5% Pred. No. 0.79;
Matches 61; Conservative 30; Mismatches 89; Indels 80; Gaps 17;

Qy	202	DSPTLQAPRERASEPKH-KNQAE	223
D _b	500	PTPDPEGDRVRAEMPQNLRKOLE	522

```

RESULT 15
US-08-916-917-14
; Sequence 14, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip Thomas
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3', KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,917
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8549-0006-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1101 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5856132e
; US-08-916-917-14

```

Query Match 6.0%; Score 73; DB 2; Length 1101;
Best Local Similarity 21.7%; Pred. NO. 16;
Matches 44; Conservative 31; Mismatches 84; Indels 44; Gaps 11

Qy	114	LVTDKGNADLKKKIQOQELSAQAP-----SPPAHS-GFHRVQRPV-----	154
Db	413	LEFSIKIDLPKGAALLNLIQYCKAPALSSKASAESSESKVRLLYYVNNLLLDHRRF	472
Qy	155	YLQSGKYIS---LLPKENKTRGSKWKMDREFLNRRHLG-EPEASTQFTQNY-----Q	201
Db	473	LLRRGEYVLMWQISGRGEGDQSGFNADKLTATNPDKENSMSISILLDNYPHIALPKHQ	532
Qy	202	DSPTLQAPRRERASEPKH-KNQAE	223
Db	533	PTDPPEGDRVRAEMPNIQRKOLE	555

Search completed: March 24, 2001, 13:23:55
Job time: 145 sec

THIS PAGE BLANK (USPTO)


```

SQ  SEQUENCE 107 AA; 11271 MH; 825BA1AEB1422390 CRC64;

Query Match 21.2%; Score 141; DB 1; Length 107;
Best Local Similarity 42.4%; Pred. NO. 4.5e-07;
Matches 25; Conservative 7; Mismatches 27; Indels 0; Gaps 0;

QY 33 LEUKESCIRNQDETCCORAPDNCSESHCAKSEGSILCOTVFFGQYRACPCLRNLTC 91
|| | | : || | | : || | | | : | | | | | |
DB 28 LEEGELCLNAQCKSGCCHHSSALSLARCAPKASENSECSPQTIYGVYKPCRGUTC 86

RESULT 3
COL_HUMAN COL_HUMAN STANDARD; PRT; 112 AA.
ID COL_HUMAN
AC P04118;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE COLIPASE PRECURSOR.
GN CLIPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RX MEDLINE=90248429; PubMed=2337598;
RA Lowe M.E., Rosenblum J.L., McEwen P., Strauss A.W.;
RT "Cloning and characterization of the human colipase cDNA.";
RL Biochemistry 29:823-828(1990).
RN [2]
RX MEDLINE=92353041; PubMed=1643046;
RA Sims H.F., Lowe M.E.;
RT "The human colipase gene: isolation, chromosomal location, and
RT tissue-specific expression.";
RL Biochemistry 31:7120-7125(1992).
RN [3]
RX MEDLINE=84104937; PubMed=6691986;
RA Sternby B., Engstroem A., Hellman U., Vihter A.M., Sternby N.-H.,
RT Borgstroem B.;
RT "The primary sequence of human pancreatic colipase.";
RL Biochim. Biophys. Acta 784:75-80(1984).
RN [4]
RX MEDLINE=93241293; PubMed=8479519;
RA van Tilbeurgh H., Egloff M.-P., Martinez C., Rugani N., Verger R.,
RT Cambillau C.;
RT "Interfacial activation of the lipase-procolipase complex by mixed
RT micelles revealed by X-ray crystallography.";
RL Nature 362:814-820(1993).
CC CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
CC CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
CC CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
CC CC INHIBITORY EFFECT ON THE LIPASE.
CC CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
CC CC SIGNAL.
CC CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; J02883; AAA52054.1; -
CC CC EMBL; M95529; AAB05618.1; -
CC CC PIR; A03163; XLHD.
CC CC PIR; A33949; A33949.
CC CC PIR; A42568; A42568.

```


DR HSSP; P02703; 1ETH.
 DR MIM; 120105; -
 DR INTERPRO; IPR001981; -
 DR PFAM; PF01114; Colipase; 1.
 DR PRINTS; PR00128; COLIPASE.
 DR PROSITE; PS00121; COLIPASE; 1.
 KW Lipid degradation; Digestion; Pancreas; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 22
 FT ENTEROSTATIN, ACTIVATION PEPTIDE
 FT (POTENTIAL).
 FT CHAIN 23 112
 FT DISULFID 34 45
 FT BY SIMILARITY.
 FT DISULFID 40 56
 FT BY SIMILARITY.
 FT DISULFID 44 78
 FT BY SIMILARITY.
 FT DISULFID 66 86
 FT BY SIMILARITY.
 FT DISULFID 80 104
 FT BY SIMILARITY.
 FT CONFLICT 68 69
 FT MISSING (IN REF. 2).
 SQ SEQUENCE 112 AA; 11954 MW; 772872EBBE7C4DF8 CRC64;
 Query Match 20.3%; Score 135.5; DB 1; Length 112;
 Best Local Similarity 28.8%; Pred. No. 1.6e-06;
 Matches 32; Conservative 17; Mismatches 51; Indels 11; Gaps 3;
 QY 12 L1FFFLFLLTRGSLPTKYNL---LELKESCIRNQDCETGCCQQRAPDNCESHCAGKSG 68
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 4 ILILLVALSVAVAPGPRGIILINLENGELCMNSACKSNCCQHSALGLARCTSMASEN 63
 QY 69 SLQQTQVFFQYRACPLRLNLTCTIYSKNEKWL-----SIAYGRCKIGRK 114
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 64 SECSVRLTYGIYKPCERGLTC---EGDKTVGSITNTNFGICHVGRSK 111
 RESULT 4
 COL2_PIG STANDARD; PRT; 95 AA.
 AC P02703;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE COLIPASE II PRECURSOR (PROCOLIPASE II).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=84104937; PubMed=6691986;
 RA Sternby B., Engstroem A., Hellman U., Vihert A.M., Sternby N.-H.,
 RA Borgstroem B.;
 RT "The primary sequence of human pancreatic colipase.";
 RL Biochim. Biophys. Acta 784:75-80(1984).
 RN [2]
 RP SEQUENCE OF 6-91.
 RX MEDLINE=74290109; PubMed=4858821;
 RA Charles M., Erlanson C., Blanchetta J.D., Joffre J., Guidoni A.A.,
 RA Rovey M.;
 RT "The primary structure of porcine colipase II. I. The amino acid
 RT sequence.";
 RL Biochim. Biophys. Acta 359:186-197(1974).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE=74290110; PubMed=4603223;
 RA Erlanson C., Charles M., Astier M., Desnuelle P.;
 RT "The primary structure of porcine colipase II. II. The disulfide
 RT bridges.";
 RL Biochim. Biophys. Acta 359:198-203(1974).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=92396238; PubMed=1522902;
 RA van Tilbeurgh H., Sarda L., Verger R., Cambillau C.;
 RT "Structure of the pancreatic lipase-procolipase complex.";
 RN Nature 359:159-162(1992).
 RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=95291181; PubMed=7773176;
 RA Egloff M.-P., Sarda L., Verger R., Cambillau C., van Tilbeurgh H.;
 RT "Crystallographic study of the structure of colipase and of the
 RT interaction with pancreatic lipase.";
 RL Protein Sci. 4:44-57(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=96279347; PubMed=8663362;
 RA Hermoso J., Pignol D., Kerfelec B., Crenon I., Chapus C.,
 RA Fontecilla-Camps J.C.;
 RT "Lipase activation by nonionic detergents. The crystal structure of
 RT the porcine lipase-colipase-tetraethylene glycol monoocetyl ether
 RT complex.";
 RL J. Biol. Chem. 271:18007-18016(1996).
 RN [7]
 RP STRUCTURE BY NMR, AND REVISION TO 37.
 RX MEDLINE=95172049; PubMed=7867624;
 RA Breg J.N., Sarda L., Cozzone P.J., Rugani N., Boelens R., Kaptein R.;
 RT "Solution structure of porcine pancreatic procolipase as determined
 RT from 1H homonuclear two-dimensional and three-dimensional NMR.";
 RL Eur. J. Biochem. 227:663-672(1995).
 CC -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
 CC THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
 CC COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
 CC INHIBITORY EFFECT ON THE LIPASE.
 CC -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
 CC SIGNAL.
 CC -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
 DR PIR; A03162; XLP62.
 DR PDB; 1PCN; 20-DEC-94.
 DR PDB; 1PCO; 20-DEC-94.
 DR PDB; 1LPA; 01-NOV-94.
 DR PDB; 1LPB; 01-NOV-94.
 DR PDB; 1ETH; 07-DEC-96.
 DR INTERPRO; IPR001981; -
 DR PFAM; PF01114; Colipase; 1.
 DR PRINTS; PR00128; COLIPASE.
 DR PROSITE; PS00121; COLIPASE; 1.
 KW Lipid degradation; Digestion; Pancreas; 3D-structure.
 FT PROPEP 1 5
 FT CHAIN 6 95
 FT SITE 6 9
 FT ENTEROSTATIN, ACTIVATION PEPTIDE.
 FT ESSENTIAL FOR THE FUNCTION OF CLP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT ESSENTIAL FOR THE FUNCTION OF CLP IN THE
 FT PRESENCE OF PHOSPHOLIPIDS.
 FT ACT_SITE 92 92
 FT DISULFID 17 28
 FT DISULFID 23 39
 FT DISULFID 27 61
 FT DISULFID 49 69
 FT DISULFID 63 87
 FT CONFLICT 37 37
 FT MISSING (IN REF. 2).
 FT CONFLICT 50 50
 FT MISSING (IN REF. 2).
 SQ SEQUENCE 95 AA; 10309 MW; 423E04549E196027 CRC64;
 Query Match 20.1%; Score 134; DB 1; Length 95;
 Best Local Similarity 34.1%; Pred. No. 1.9e-06;
 Matches 29; Conservative 11; Mismatches 37; Indels 8; Gaps 2;
 QY 33 LELKESCIRNQDCETGCCQQRAPDNCESHCAGKSGSLQQTQVFFQYRACPLRLNLTCTI 92
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 11 LDEGELCLNSAQCKSNCCQHDITLSLRCAKARENSECSAFTLYGYKPCERGLTC- 69
 QY 93 YSKNEKWL-----SIAYGRCKIGR 112
 :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | | :| | |
 Db 70 --EGDKSLVGSITNTNFGICHVGR 92
 RESULT 5
 COLB_HORSE STANDARD; PRT; 108 AA.
 ID COLB_HORSE
 AC P02705;

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROCOLIPASE B PRECURSOR (FRAGMENT)
OS Equus caballus (Horse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RN SEQUENCE FROM N.A.
RN MEDLINE=94325330; PubMed=8049250;
RN Crenon I., Granon S., Chapus C., Kerfelec B.;
RN "Molecular cloning and expression of two horse pancreatic CDNA
RN encoding colipase A and B";
RN Biochim. Biophys. Acta 1213:357-360(1994).
RN [2]
RN SEQUENCE OF 14-108.
RN MEDLINE=82046794; PubMed=7295770;
RN Bonicel J.J., Couchoud P.M., Foglizzo E., Desmuelle P., Chapus C.;
RN "Amino acid sequence of horse colipase B";
RN Biochim. Biophys. Acta 669:39-45(1981).
RN [3]
RN SEQUENCE OF 14-68.
RN MEDLINE=81021166; PubMed=7417313;
RN Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
RN "Evidence for the existence of two isocolipases in horse pancreas";
RN Biochem. Biophys. Res. Commun. 95:1245-1252(1980).
RN -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
RN THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
RN COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
RN INHIBITORY EFFECT ON THE LIPASE.
RN -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
RN SIGNAL.
RN -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN EMBL; X74344; CAA52391.1;
RN PIR; A03165; XLHOB.
RN HSSP; P02703; IETH.
RN INTERPRO; IPR001981;
RN PFAM; PF01114; Colipase; 1.
RN PROSITE; PS00121; COLIPASE; 1.
RN Lipid degradation; Digestion; Pancreas; Signal.
RN NON_TER 1
RN SIGNAL 1
RN PROPEP 13
RN CHAIN 18
RN BINDING 19
RN BY SIMILARITY 65
RN DISULFID 30
RN BY SIMILARITY 41
RN DISULFID 36
RN BY SIMILARITY 52
RN DISULFID 40
RN BY SIMILARITY 74
RN DISULFID 62
RN BY SIMILARITY 82
RN DISULFID 76
RN BY SIMILARITY 100
RN CONFLICT 35
RN CONFLICT 35
RN CONFLICT 42
RN CONFLICT 42
RN CONFLICT 108
RN CONFLICT 108
RN SEQUENCE 108 AA; 11618 MW; 3747F49ABA08B30 CRC64;
Query Match 20.1%; Score 134; DB 1; Length 108;
Best Local Similarity 39.7%; Pred. No. 2.1e-06;
Matches 25; Conservative 6; Mismatches 32; Indels 0; Gaps 0;
OY 33 LEKESICRNQDCGCGCORAFDNCSEHCAEKGSGSLCQTVFFGQYRACPCRLNLCI 92
DB 24 LEAGEICMNSAQCKBCCCHRESSLSLARCAAKASECSAWTLYGVYKPCERGLTCQ 83

OY 93 YSK 95
DB 84 VDK 86
RESULT 6
COLA_HORSE STANDARD; PRT; 106 AA.
ID P02704;
AC 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PROCOLIPASE A PRECURSOR (FRAGMENT)
OS Equus caballus (Horse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RN SEQUENCE FROM N.A.
RN TISSUE=PANCREAS;
RN MEDLINE=94325330; PubMed=8049250;
RN Crenon I., Granon S., Chapus C., Kerfelec B.;
RN "Molecular cloning and expression of two horse pancreatic CDNA
RN encoding colipase A and B";
RN Biochim. Biophys. Acta 1213:357-360(1994).
RN [2]
RN SEQUENCE OF 12-106.
RN MEDLINE=84104937; PubMed=6691986;
RN Sternby B., Engstroem A., Hellman U., Viherth A.M., Sternby N.-H.,
RN Borgstroem B.;
RN "The primary sequence of human pancreatic colipase";
RN Biochim. Biophys. Acta 784:75-80(1984).
RN [3]
RN SEQUENCE OF 12-106.
RN MEDLINE=82186702; PubMed=7075593;
RN Pierrot M., Astier J.-P., Astier M., Charles M., Drenth J.;
RN "Pancreatic colipase: crystallographic and biochemical aspects";
RN Eur. J. Biochem. 123:347-354(1982).
RN [4]
RN SEQUENCE OF 12-66.
RN MEDLINE=81021166; PubMed=7417313;
RN Julien R., Bechis G., Gregoire J., Rathelot J., Rochat H., Sarda L.;
RN "Evidence for the existence of two isocolipases in horse pancreas";
RN Biochem. Biophys. Res. Commun. 95:1245-1252(1980).
RN -1- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS
RN THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT
RN COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN
RN INHIBITORY EFFECT ON THE LIPASE.
RN -1- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY
RN SIGNAL.
RN -1- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPASE.
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
RN or send an email to license@isb-sib.ch).
RN EMBL; X74503; CAA52611.1;
RN PIR; A03164; XLHOA.
RN HSSP; P02703; IETH.
RN INTERPRO; IPR001981;
RN PFAM; PF01114; Colipase; 1.
RN PROSITE; PS00121; COLIPASE; 1.
RN Lipid degradation; Digestion; Pancreas; Signal.
RN NON_TER 1
RN SIGNAL 1
RN PROPEP 11
RN CHAIN 12
RN BY SIMILARITY 16
RN DISULFID 17
RN BY SIMILARITY 106
RN DISULFID 28
RN BY SIMILARITY 39
RN DISULFID 34
RN BY SIMILARITY 50
ENTEROSTATIN, ACTIVATION PEPTIDE.
COLIPASE A.
BY SIMILARITY.
BY SIMILARITY.

FT	CHAIN	23	111	COLIPASE.	
FT	DISULFID	34	45	BY SIMILARITY.	
FT	DISULFID	40	56	BY SIMILARITY.	
FT	DISULFID	44	78	BY SIMILARITY.	
FT	DISULFID	66	86	BY SIMILARITY.	
FT	DISULFID	80	104	BY SIMILARITY.	
SEQUENCE	111 AA; 11899 MW; 7FF37DE/CI69616B CRC64;				
Query Match					
Best Local Similarity		35.98;	Pred. No. 4.1e-06;		
Matches 28; Conservative 10; Mismatches 38; Indels 2; Gaps					
QY	33	LELKESCIRNQDCTGCCQORAPDNCESHCAEKSEGSICQTQVFQYRACPLRLNLTGI 92			
		1:	1:	1:	1:
		1:	1:	1:	1:
Db	28	LNGELCLNSAQCKSQCCQHDPSPLGLARCADKARENSGCSPTQIYGIYILCPCERGLTCD 87			
QY	93	YSKN--EKWLSIAYGRQC 108			
		1:	1:	1:	1:
Db	88	GDKSIIGAITNTNYGICQ 105			
RESULT 8					
ID	COL_RAT	STANDARD;	PRT;	112 AA.	
AC	P17084;				
DT	01-AUG-1990 (Rel. 15, Created)				
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	COLIPASE PRECURSOR.				
GN	CLPS.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=90179738; PubMed=2129524;				
RT	Wicker C., Puigserver A.;				
RT	"Rat pancreatic colipase mRNA: nucleotide sequence of a cDNA clone				
RT	and nutritional regulation by a lipidic diet."				
RL	Biochem. Biophys. Res. Commun. 167:130-136(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=94262798; PubMed=8203536;				
RT	Payne R.M., Sims H.F., Jennens M.L., Lowe M.E.;				
RT	"Rat pancreatic lipase and two related proteins: enzymatic properties				
RT	and mRNA expression during development."				
CC	Am. J. Physiol. 266:G914-G921(1994).				
CC	-!- FUNCTION: COLIPASE IS A COFACTOR OF PANCREATIC LIPASE. IT ALLOWS				
CC	THE LIPASE TO ANCHOR ITSELF TO THE LIPID-WATER INTERFACE. WITHOUT				
CC	COLIPASE THE ENZYME IS WASHED OFF BY BILE SALTS, WHICH HAVE AN				
CC	INHIBITORY EFFECT ON THE LIPASE.				
CC	-!- FUNCTION: ENTEROSTATIN HAS A BIOLOGICAL ACTIVITY AS A SATIETY				
CC	SIGNAL.				
CC	-!- SUBUNIT: FORM A 1:1 STOICHIOMETRIC COMPLEX WITH PANCREATIC LIPAS				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collabora				
CC	between the Swiss Institute of Bioinformatics and the EMBL outsta				
CC	the European Bioinformatics Institute. There are no restrictions o				
CC	use by non-profit institutions as long as its content is in n				
CC	modified and this statement is not removed. Usage by and for comm				
CC	entities requires a license agreement (See http://www.isb-sib.ch/ann				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M33333; AAA40943.1; -				
DR	EMBL; M58370; AAA20505.1; -				
DR	PIR; A34623; A34623.				
DR	HSP; P02703; 1ETH.				
DR	INTERPRO: IPR001981; -				
DR	PFAM; PF01114; Colipase.1.				
DR	PRINTS; PS00128; COLIPASE.				
DR	PROSITE; PS00121; COLIPASE; 1.				
DR	Lipid degradation; Digestion; Pancreas; Signal.				
KW					

SIGNAL	1	17	ENTEROSTATIN, ACTIVATION PEPTIDE
PROPEP	18	22	(POTENTIAL).
CHAIN	23	112	COLIPASE.
DISULFID	34	45	BY SIMILARITY.
DISULFID	40	56	BY SIMILARITY.
DISULFID	44	78	BY SIMILARITY.
DISULFID	66	86	BY SIMILARITY.
DISULFID	80	104	BY SIMILARITY.
CONFLICT	18	18	V -> A. (IN REF. 2).
SEQUENCE	112 AA;	1228 MW;	FB5659DA59D049D CRC64;
Query Match	18.3%;	Score 122;	DB 1; Length 112;
Best Local Similarity	31.9%;	Pred. No. 3e-05;	
Matches	29;	Conservative	8; Mismatches 38; Indels 16; Gaps 2;
QY	33	LELKESCIRNPQCTCCORAPDNCNESHCAEKGSGSLCQTQVFFGQYRACPCLRLTTC-	91
DBb	28	LEDGEICVNSMOCKSRCCOHDHILGIARTKHAMENSECSPTLLGYIYRCPCEGLTCE	87
QY	92	-----LYSKNEKWLSTAYGRCQKIGROK	114
DBb	88	GDRSIGAITNTN-----YGVCLDSTRSK	111
RESULT	9		
PDGA_XENLA			
ID	PDGA_XENLA	STANDARD;	PRT; 226 AA.
AC	PL3698;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	01-OCT-2000 (Rel. 40, Last annotation update)		
DE	PLATELET-DERIVED GROWTH FACTOR, A CHAIN PRECURSOR (PDGF A-CHAIN)		
OS	(PDGFA).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;		
OC	Xenopodinae; Xenopus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OOCYTE;		
RC	MEDLINE=88321676; PubMed=3413486;		
RA	Mercola M., Melton D.A., Stiles C.D.;		
RT	"platelet-derived growth factor A chain is maternally encoded in		
RT	Xenopus embryos.;"		
RL	Science 241:1223-1225(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OOCYTE;		
RC	MEDLINE=90175018; PubMed=2308861;		
RA	Bejcek B.E., Li D.Y., Deuel T.F.;		
RT	"Nucleotide sequence of a cDNA clone of Xenopus platelet-derived		
RT	growth factor A-chain.;"		
RL	Nucleic Acids Res. 18:680-680(1990).		
CC	-1- FUNCTION: PLATELET-DERIVED GROWTH FACTOR IS A POTENT MITOGEN FOR		
CC	CELLS OF MESENCHYMAL ORIGIN. BINDING OF THIS GROWTH FACTOR TO ITS		
CC	AFFINITY RECEPTOR ELICITS A VARIETY OF CELLULAR RESPONSES. IT IS		
CC	RELEASED BY PLATELETS UPON WOUNDING AND PLAYS AN IMPORTANT ROLE		
CC	IN STIMULATING ADVANCE CELLS TO GROW AND THEREBY HEAL THE WOUND.		
CC	-1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED DIMER OF NONIDENTICAL (A		
CC	AND B) CHAINS. HOMODIMERS OF A AND B CHAINS ARE IMPLICATED IN		
CC	TRANSFORMATION PROCESSES.		
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A		
CC	SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE LONG FORM		
CC	CONTAINS A BASIC INSERT WHICH ACTS AS A CELL RETENTION SIGNAL.		
CC	-1- MISCELLANEOUS: A-A AND B-B, AS WELL AS A-B, DIMERS CAN BIND TO THE		
CC	PDGF RECEPTOR.		
CC	-1- SIMILARITY: BELONGS TO THE PDGF/VEGF FAMILY OF GROWTH FACTORS.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on its		

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
 CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
 CC ADULT.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57405; CAA040667.1; -
 CC HSSP: P00740; 11XA.
 CC INTERPRO: IPR000152; -
 CC INTERPRO: IPR000561; -
 CC INTERPRO: IPR000800; -
 CC INTERPRO: IPR001438; -
 CC INTERPRO: IPR001881; -
 CC INTERPRO: IPR002049; -
 CC INTERPRO: IPR002110; -
 CC PFAM: PF00008; EGF_36.
 CC PFAM: PF00023; ank; 6.
 CC PFAM: PF00066; notch; 3.
 CC PRINTS: PR00010; EGFBL00D.
 CC PRINTS: PR00011; EGFAMININ.
 CC PROSITE: PS00088; ANK_REPEAT; 4.
 CC PROSITE: PS0297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC PROSITE: PS00022; EGF_1; 35.
 CC PROSITE: PS01186; EGF_2; 26.
 CC PROSITE: PS01187; EGF_CA; 21.
 CC Differentiation: Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
 FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1724 1746 POTENTIAL.
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 58 EGF-LIKE 1.
 FT DOMAIN 59 99 EGF-LIKE 2.
 FT DOMAIN 102 139 EGF-LIKE 3.
 FT DOMAIN 140 176 EGF-LIKE 4.
 FT DOMAIN 178 216 EGF-LIKE 5.
 FT DOMAIN 218 255 EGF-LIKE 6.
 FT DOMAIN 257 293 EGF-LIKE 7.
 FT DOMAIN 295 333 EGF-LIKE 8.
 FT DOMAIN 335 371 EGF-LIKE 9.
 FT DOMAIN 372 410 EGF-LIKE 10.
 FT DOMAIN 412 450 EGF-LIKE 11.
 FT DOMAIN 452 488 EGF-LIKE 12.
 FT DOMAIN 490 526 EGF-LIKE 13.
 FT DOMAIN 528 564 EGF-LIKE 14.
 FT DOMAIN 566 601 EGF-LIKE 15.
 FT DOMAIN 603 639 EGF-LIKE 16.
 FT DOMAIN 641 676 EGF-LIKE 17.
 FT DOMAIN 678 714 EGF-LIKE 18.
 FT DOMAIN 716 751 EGF-LIKE 19.
 FT DOMAIN 753 789 EGF-LIKE 20.
 FT DOMAIN 791 827 EGF-LIKE 21.
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23.
 FT DOMAIN 907 943 EGF-LIKE 24.
 FT DOMAIN 945 981 EGF-LIKE 25.
 FT DOMAIN 983 1019 EGF-LIKE 26.
 FT DOMAIN 1021 1057 EGF-LIKE 27.
 FT DOMAIN 1059 1095 EGF-LIKE 28.
 FT DOMAIN 1097 1143 EGF-LIKE 29.

FT DOMAIN 1145 1181 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1183 1219 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1221 1265 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1267 1305 EGF-LIKE 33.
 FT DOMAIN 1307 1346 EGF-LIKE 34.
 FT DOMAIN 1348 1384 EGF-LIKE 35.
 FT DOMAIN 1387 1426 EGF-LIKE 36.
 FT DOMAIN 1449 1462 CYS-RICH.
 FT DOMAIN 1865 2076 6 X ANK MOTIF REPEATS.
 FT REPEAT 1865 1910 ANK MOTIF 1.
 FT REPEAT 1912 1942 ANK MOTIF 2.
 FT REPEAT 1944 1975 ANK MOTIF 3.
 FT REPEAT 1978 2009 ANK MOTIF 4.
 FT REPEAT 2011 2042 ANK MOTIF 5.
 FT REPEAT 2044 2076 ANK MOTIF 6.
 FT DISULFID 24 37 BY SIMILARITY.
 FT DISULFID 31 46 BY SIMILARITY.
 FT DISULFID 48 57 BY SIMILARITY.
 FT DISULFID 63 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 127 BY SIMILARITY.
 FT DISULFID 129 138 BY SIMILARITY.
 FT DISULFID 144 155 BY SIMILARITY.
 FT DISULFID 149 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 182 195 BY SIMILARITY.
 FT DISULFID 189 204 BY SIMILARITY.
 FT DISULFID 206 215 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.
 FT DISULFID 266 281 BY SIMILARITY.
 FT DISULFID 283 292 BY SIMILARITY.
 FT DISULFID 299 312 BY SIMILARITY.
 FT DISULFID 306 321 BY SIMILARITY.
 FT DISULFID 323 332 BY SIMILARITY.
 FT DISULFID 339 350 BY SIMILARITY.
 FT DISULFID 344 359 BY SIMILARITY.
 FT DISULFID 361 370 BY SIMILARITY.
 FT DISULFID 376 387 BY SIMILARITY.
 FT DISULFID 381 398 BY SIMILARITY.
 FT DISULFID 400 409 BY SIMILARITY.
 FT DISULFID 416 429 BY SIMILARITY.
 FT DISULFID 423 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 456 467 BY SIMILARITY.
 FT DISULFID 461 476 BY SIMILARITY.
 FT DISULFID 478 487 BY SIMILARITY.
 FT DISULFID 494 505 BY SIMILARITY.
 FT DISULFID 499 514 BY SIMILARITY.
 FT DISULFID 516 525 BY SIMILARITY.
 FT DISULFID 532 543 BY SIMILARITY.
 FT DISULFID 537 552 BY SIMILARITY.
 FT DISULFID 554 563 BY SIMILARITY.
 FT DISULFID 570 580 BY SIMILARITY.
 FT DISULFID 575 589 BY SIMILARITY.
 FT DISULFID 591 600 BY SIMILARITY.
 FT DISULFID 607 618 BY SIMILARITY.
 FT DISULFID 612 627 BY SIMILARITY.
 FT DISULFID 629 638 BY SIMILARITY.
 FT DISULFID 645 655 BY SIMILARITY.
 FT DISULFID 650 664 BY SIMILARITY.
 FT DISULFID 666 675 BY SIMILARITY.
 FT DISULFID 682 693 BY SIMILARITY.
 FT DISULFID 687 702 BY SIMILARITY.
 FT DISULFID 704 713 BY SIMILARITY.
 FT DISULFID 720 730 BY SIMILARITY.
 FT DISULFID 725 739 BY SIMILARITY.
 FT DISULFID 741 750 BY SIMILARITY.
 FT DISULFID 757 768 BY SIMILARITY.

LE
LE
LE
LE
LE
LE

FT FT FT FT FT FT FT

LE
LE
LE
LE
LE

LE
LE
LE
LE

4 F F F F

LJ
LJ
LJ
LJ

Q B M

Qy Db

Qy Db

RES
RSI
ID

ID AC DT DT

DT DE GN OS

$$\begin{array}{cccc} \text{OC} & \text{OC} & \text{OC} & \text{RN} \end{array}$$

RP RC RX RA

RT RT RT RT RT

22222

UUUU

•

CC EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -.
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR HSSP; P00740; lixa.
DR FLYBASE; FBgn0000368; crb.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR001438; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF; 27.
DR PFAM; PF00054; laminin_G; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX HYDROXYL; 15.
DR PROSITE; PS00022; EGF_1; 26.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS01187; EGF_CA; 15.
KW Differentiation; Repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal; Phosphorylation.
FT SIGNAL 1 90
FT CHAIN 91 2139 CROMES PROTEIN.
FT DOMAIN 91 2084 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2085 2111 POTENTIAL.
FT DOMAIN 2112 2139 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 267 303 EGF-LIKE 1.
FT DOMAIN 306 343 EGF-LIKE 2.
FT DOMAIN 348 386 EGF-LIKE 3.
FT DOMAIN 388 425 EGF-LIKE 4.
FT DOMAIN 427 463 EGF-LIKE 5.
FT DOMAIN 464 500 EGF-LIKE 6.
FT DOMAIN 501 532 EGF-LIKE 7.
FT DOMAIN 545 581 EGF-LIKE 8.
FT DOMAIN 582 611 EGF-LIKE 9.
FT DOMAIN 609 646 EGF-LIKE 10.
FT DOMAIN 648 685 EGF-LIKE 11.
FT DOMAIN 687 723 EGF-LIKE 12.
FT DOMAIN 725 761 EGF-LIKE 13.
FT DOMAIN 763 800 EGF-LIKE 14.
FT DOMAIN 802 838 EGF-LIKE 15.
FT DOMAIN 840 902 EGF-LIKE 16.
FT DOMAIN 904 948 EGF-LIKE 17.
FT DOMAIN 942 978 EGF-LIKE 18.
FT DOMAIN 980 1021 EGF-LIKE 19.
FT DOMAIN 1027 1243 EGF-LIKE 20.
FT DOMAIN 1481 1517 EGF-LIKE 21.
FT DOMAIN 1759 1795 EGF-LIKE 22.
FT DOMAIN 1797 1833 EGF-LIKE 23.
FT DOMAIN 1835 1871 EGF-LIKE 24.
FT DOMAIN 1874 1915 EGF-LIKE 25.
FT DOMAIN 1915 1951 EGF-LIKE 26.
FT DOMAIN 1953 1989 EGF-LIKE 27.
FT DOMAIN 1991 2029 EGF-LIKE 28.
FT DOMAIN 2030 2070 EGF-LIKE 29.
FT DISULFID 271 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 293 302 BY SIMILARITY.
FT DISULFID 310 321 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 333 342 BY SIMILARITY.
FT DISULFID 352 363 BY SIMILARITY.
FT DISULFID 357 374 BY SIMILARITY.
FT DISULFID 376 385 BY SIMILARITY.
FT DISULFID 392 403 BY SIMILARITY.
FT DISULFID 397 412 BY SIMILARITY.
FT DISULFID 414 424 BY SIMILARITY.
FT DISULFID 431 442 BY SIMILARITY.
FT DISULFID 436 451 BY SIMILARITY.
FT DISULFID 453 462 BY SIMILARITY.
FT DISULFID 468 479 BY SIMILARITY.
FT DISULFID 473 488 BY SIMILARITY.
FT DISULFID 490 499 BY SIMILARITY.
FT DISULFID 505 515 BY SIMILARITY.
FT DISULFID 509 520 BY SIMILARITY.
FT DISULFID 522 531 BY SIMILARITY.
FT DISULFID 549 562 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 586 597 BY SIMILARITY.
FT DISULFID 591 602 BY SIMILARITY.
FT DISULFID 604 610 BY SIMILARITY.
FT DISULFID 613 624 BY SIMILARITY.
FT DISULFID 618 634 BY SIMILARITY.
FT DISULFID 636 645 BY SIMILARITY.
FT DISULFID 652 664 BY SIMILARITY.
FT DISULFID 659 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT DISULFID 691 702 BY SIMILARITY.
FT DISULFID 696 711 BY SIMILARITY.
FT DISULFID 713 722 BY SIMILARITY.
FT DISULFID 729 740 BY SIMILARITY.
FT DISULFID 734 749 BY SIMILARITY.
FT DISULFID 751 760 BY SIMILARITY.
FT DISULFID 767 778 BY SIMILARITY.
FT DISULFID 772 787 BY SIMILARITY.
FT DISULFID 789 799 BY SIMILARITY.
FT DISULFID 806 817 BY SIMILARITY.
FT DISULFID 811 826 BY SIMILARITY.
FT DISULFID 828 837 BY SIMILARITY.
FT DISULFID 844 855 BY SIMILARITY.
FT DISULFID 849 890 BY SIMILARITY.
FT DISULFID 892 901 BY SIMILARITY.
FT DISULFID 908 919 BY SIMILARITY.
FT DISULFID 913 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 946 957 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 968 977 BY SIMILARITY.
FT DISULFID 984 995 BY SIMILARITY.
FT DISULFID 989 1009 BY SIMILARITY.
FT DISULFID 1011 1020 BY SIMILARITY.
FT DISULFID 1211 1222 BY SIMILARITY.
FT DISULFID 1216 1231 BY SIMILARITY.
FT DISULFID 1233 1242 BY SIMILARITY.
FT DISULFID 1485 1496 BY SIMILARITY.
FT DISULFID 1490 1505 BY SIMILARITY.
FT DISULFID 1507 1516 BY SIMILARITY.
FT DISULFID 1763 1774 BY SIMILARITY.
FT DISULFID 1768 1783 BY SIMILARITY.
FT DISULFID 1785 1794 BY SIMILARITY.
FT DISULFID 1801 1812 BY SIMILARITY.
FT DISULFID 1806 1821 BY SIMILARITY.
FT DISULFID 1823 1832 BY SIMILARITY.
FT DISULFID 1839 1850 BY SIMILARITY.
FT DISULFID 1844 1859 BY SIMILARITY.
FT DISULFID 1861 1870 BY SIMILARITY.
FT DISULFID 1878 1889 BY SIMILARITY.
FT DISULFID 1883 1903 BY SIMILARITY.
FT DISULFID 1905 1914 BY SIMILARITY.
FT DISULFID 1919 1930 BY SIMILARITY.
FT DISULFID 1924 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1957 1968 BY SIMILARITY.
FT DISULFID 1962 1977 BY SIMILARITY.
FT DISULFID 1979 1988 BY SIMILARITY.
FT DISULFID 1995 2008 BY SIMILARITY.
FT DISULFID 2002 2017 BY SIMILARITY.
FT DISULFID 2019 2028 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).

RX MEDLINE=93048835; PubMed=1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -I- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -I- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -I- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -I- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -I- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z11886; CAA77941.1; -

-1- FUNCTION: MAY HAVE A ROLE IN NEUROENDOCRINE DIFFERENTIATION.
-1- SUBUNIT: MONOMER.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
-1- TISSUE SPECIFICITY: FAI IS FOUND WITHIN THE STROMAL CELLS IN CLOSE CONTACT TO THE VASCULAR STRUCTURE OF PLACENTAL VILLI, YOLK SAC, FETAL LIVER, ADRENAL CORTEX AND PANCREAS AND IN THE BETA CELLS OF THE ISLETS OF LANGERHANS IN THE ADULT PANCREAS. FOUND ALSO IN SOME FORMS OF NEUROENDOCRINE LUNG TUMOR TISSUE.
-1- PPM: N- AND O-GLYCOSYLATED.
-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z12172; CAA78163.1; -
EMBL; U15979; AAA75364.1; -
EMBL; U15981; AAA75365.1; -
HSSP; P00750; ITPG.
MIM; 176290; -
INTERPRO; IPR000152; -
INTERPRO; IPR000561; -
PFAM; PF00008; EGF; 6.
PROSITE; PS00010; ASX_HYDROXYL; 1.
PROSITE; PS00022; EGF_1; 5.
PROSITE; PS01186; EGF_2; 5.
Antigen; Transmembrane; EGF-like domain; Repeat; Glycoprotein; Signal;
Alternative splicing.
FT SIGNAL 1 23
FT CHAIN 24 383
FT CHAIN 24 303
FT CHAIN 24 303
FT DOMAIN 24 303
FT DOMAIN 304 327
FT TRANSMEM 328 383
FT DOMAIN 24 55
FT DOMAIN 53 86
FT DOMAIN 88 125
FT DOMAIN 127 168
FT DOMAIN 170 206
FT DOMAIN 208 245
FT DOMAIN 26 37
FT DISULFID 30 43
FT DISULFID 45 54
FT DISULFID 57 68
FT DISULFID 63 74
FT DISULFID 76 85
FT DISULFID 92 103
FT DISULFID 97 113
FT DISULFID 115 124
FT DISULFID 131 144
FT DISULFID 138 156
FT DISULFID 158 167
FT DISULFID 174 185
FT DISULFID 179 194
FT DISULFID 196 205
FT DISULFID 212 223
FT DISULFID 217 233
FT DISULFID 235 244
FT CARBOHYD 94 94
FT CARBOHYD 100 100
FT CARBOHYD 143 143
FT CARBOHYD 163 163
FT CARBOHYD 165 165
FT CARBOHYD 172 172
FT CARBOHYD 214 214
FT CARBOHYD 222 222
FT CARBOHYD 251 251

N-LINKED (GLCNAC. . .).
PARTIAL.
N-LINKED (GLCNAC. . .) (PARTIAL).
N-LINKED (GLCNAC. . .) (PARTIAL).
PARTIAL.
PARTIAL.

FT CARBOHYD 260 260 PARTIAL.
FT VARSPLIC 229 301 MISSING (IN SHORT ISOFORM).
FT VARIANT 347 347 MISSING (IN CLONE HDLKAAG).
FT /FTID-VAR_002274. 1).
FT CONFLICT 46 47 QP -> HV (IN REF. 1).
FT CONFLICT 108 108 G -> D (IN REF. 3).
SQ SEQUENCE 383 AA; 41143 MW; 2C55C6E355B4415B CRC64;

Query Match 11.3%; Score 75; DB 1; Length 383;
Best Local Similarity 34.4%; Pred. No. 2.4;
Matches 21; Conservative 6; Mismatches 22; Indels 12; Gaps 5;

QY 35 LKESCI RNODCTGCGCORAPDNCHSAEKSEGS LCOFQVFFGVGRAC---PCLRNLTG 91
DB 53 LCDQCVTSPGCLHGLGGE-PGQC---ICTD-GWDGELCDRDV-----RACSSAPCANNGTC 103
QY 92 I 92
DB 104 V 104

RESULT 15
DISJL_BOTJA STANDARD; PRT; 571 AA.
AC P30431;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE VENOM METALLOPROTEINASE JARAHAGIN PRECURSOR (EC 3.4.24.73)
DE (HF2-PROTEINASE) [CONTAINS: DISINTEGRIN] (FRAGMENT).
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OC [1]
RN SEQUENCE FROM N.A.
RP TISSUE=VENOM GLAND;
RC MEDLINE=93054601; PubMed=1385408;
RX Paine M.J.I., Desmond H.P., Theakston R.D.G., Crampton J.M.;
RA "Purification, cloning, and molecular characterization of a high
RT molecular weight hemorrhagic metalloprotease, jarahagin, from
RT Bothrops jararaca venom. Insights into the disintegrin gene family.";
RL J. Biol. Chem. 267:22869-22876(1992).
CC -1- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC ACTS IN HEMORRHAGE (BY SIMILARITY).
CC -1- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC AND COLLAGEN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF 10-HIS-I-LEU-11, 14-ALA-I-LEU-15,
CC 16-TYR-I-LEU-17 AND 24-PHE-I-PHE-25 BONDS IN INSULIN B CHAIN.
CC -1- COFACTOR: BINDS ONE ZINC ION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPLYLYSIN SUBFAMILY.
CC -1- SIMILARITY: IN THE CENTRAL REGION; BELONGS TO THE DISINTEGRIN
CC FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X68251; CAA48323.1; -
PIR; S24789; S24789.
PIR; A44463; A44463.
HSSP; P17494; 1KST.
MEROPS; M12.136; -
INTERPRO; IPR000130; -

```

DR INTERPRO; IPR001590; -
DR INTERPRO; IPR001762; -
DR INTERPRO; IPR002870; -
DR PFAM; PF01562; Pep_M12B_propep; 1.
DR PFAM; PF01421; Reptolysin; 1.
DR PFAM; PF00200; disintegrin; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS00214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW Cell adhesion; Venom; Zymogen.
FT NON_TER 1 1
FT PROPEP <1 151 POTENTIAL.
FT CHAIN 152 355 JARAHAGIN.
FT CHAIN 356 447 DISINTEGRIN.
FT PROPEP 448 571
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT METAL 295 295 ZINC (CATALYTIC) (PROBABLE).
FT ACT_SITE 296 296 BY SIMILARITY.
FT METAL 299 299 ZINC (CATALYTIC) (PROBABLE).
FT METAL 305 305 ZINC (CATALYTIC) (PROBABLE).
FT DISULFID 270 350 BY SIMILARITY.
FT DISULFID 310 317 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 379 390 BY SIMILARITY.
FT DISULFID 389 412 BY SIMILARITY.
FT DISULFID 403 409 BY SIMILARITY.
FT DISULFID 408 434 BY SIMILARITY.
FT DISULFID 421 441 BY SIMILARITY.
FT CARBOHYD 333 333 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 571 AA; 63982 MW; F5E549F8BF61177B CRC64;

```

```

Query Match      11.2%; Score 74.5; DB 1; Length 571;
Best Local Similarity 26.5%; Pred. No. 3.8;
Matches 30; Conservative 7; Mismatches 29; Indels 47; Gaps 7;

QY 32 LLELKESCI--RNQDCETGCCORAP-----DNCE----- 58
Db 370 LLEVGECDGTPENCQNECCDAATCKLKSQGGHGDCCCEQCKFKSGTGTCSRSMSECD 429
QY 59 --SHCAEKSGSEGLCOPTQVFF--GQYRACPLRLNLTCTIYSKNEKWLSTIAYGRC 107
Db 430 PAEHC TGQSSE---CPADVFHKNGQ----PCLDNYGYCYNGN---CPIMVHQC 472

```

Search completed: March 24, 2001, 13:24:24
Job time: 170 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:26 ; Search time 47.63 Seconds
(without alignments)
86.866 Million cell updates/sec

Title: US-09-215-435-179
Perfect score: 666
Sequence: 1 MMLPQWLLLLFLFFFL.....IAYGRCKIGROKLAKMFF 121

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	666	100.0	121	Y35930	Extended human sec
2	666	100.0	122	Y91513	Human secreted pro
3	619	92.9	120	Y12687	Human 5' EST secre
4	85.5	12.8	271	Y44934	Human dickkopf-1 h
5	82.5	12.4	179	W73019	Human cysteine-ric
6	82.5	12.4	207	Y92074	Human DKR-2 splice
7	82.5	12.4	259	Y93360	Human PRO1316 (UNO
8	82.5	12.4	259	Y92072	Murine DKR-2. Mus
9	82.5	12.4	259	Y92073	Murine DKR-2. Homo
10	77	11.6	125	R25449	PDGF analogue #9.
11	77	11.6	125	R87521	Glioma cell PDGF A
12	76	11.4	802	R44242	F-spondin of chick

13	75	11.3	383	15	R56166	Neuroendocrine tum
14	75	11.3	1148	20	W87895	Human JAGGED2 prot
15	74	11.1	385	15	R56167	Neuroendocrine tum
16	74	11.1	807	20	Y04262	Human vascular smo
17	73	11.0	125	14	R44590	Monomeric PDGF-A.
18	73	11.0	807	14	R44241	F-spondin (FP5-9).
19	73	11.0	807	20	Y04263	Bovine vascular sm
20	73	11.0	807	20	Y04264	Rat vascular smoot
21	72.5	10.9	266	19	W73018	Human cysteine-ric
22	72.5	10.9	266	20	Y41757	Human PRO1008 prot
23	72.5	10.9	266	21	Y92071	Human DKR-1. Homo
24	72	10.8	125	14	R44584	Monomeric PDGF-A.
25	72	10.8	126	12	R12880	rPDGF A1 from pCFM
26	72	10.8	126	13	R25431	Glial PDGF A cha
27	72	10.8	211	12	R15645	PDGF D-1 A-chain p
28	72	10.8	211	14	R38918	Human Platelet Der
29	72	10.8	211	16	R84759	Human platelet-der
30	70.5	10.6	316	21	Y44862	Human taste cell s
31	70.5	10.6	1055	19	W44298	Human serrate 2 pr
32	70.5	10.6	1212	19	W44299	Human serrate 2.
33	70.5	10.6	1257	17	W05834	Human Serrate-2 (H
34	70.5	10.6	1257	21	Y59598	Human Serrate prot
35	69.5	10.4	224	19	W73017	Human cysteine-ric
36	69.5	10.4	224	21	Y92075	Human DKR-4. Homo
37	69.5	10.4	568	14	R44243	Xenopus F-spondin.
38	69.5	10.4	737	11	R05270	Amino acid sequenc
39	69	10.4	299	21	Y76014	Murine EGF family
40	69	10.4	299	21	Y76077	Murine EGF family
41	69	10.4	613	19	W73009	Cobra venom protea
42	69	10.4	621	19	W73013	Cobra venom moca
43	69	10.4	921	21	Y32297	Corn polycomb prot
44	68.5	10.3	79	17	R91704	NamNAP. Necator a
45	68.5	10.3	79	20	Y30412	Nematode extracted

ALIGNMENTS

RESULT 1
Y35930
ID Y35930 standard; Protein; 121 AA.
XX
AC Y35930;
XX
DT 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein sequence, SEQ ID NO. 179.
XX

Secreted protein; human; cytokine; cellular proliferation; cell movement;
cellular differentiation; immune system regulator; anti-inflammatory;
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
genetic disease.
XX Homo sapiens.
XX OS
XX WO9931236-A2.
XX PN
XX PD 24-JUN-1999.
XX PF 17-DEC-1998; 98WO-IB02122.
XX PR 10-AUG-1998; 98US-0096116.
XX PR 17-DEC-1997; 97US-0069957.
XX PR 09-FEB-1998; 98US-0074121.
XX PR 13-APR-1998; 98US-0081563.
XX PA (GEST) GENSET.
XX FI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX WPI; 1999-385906/32.
XX DR N-PSDB; X97614.

XX New isolated human secreted proteins

XX PS Claim 9; Page 212; 516pp; English.

XX This sequence is encoded by an extended human secreted protein coding

CC sequence of the invention. The secreted proteins can be used in treating

CC or controlling a variety of human conditions. The secreted proteins may

CC act as cytokines or may affect cellular proliferation or differentiation

CC or may act as immune system regulators, haematopoiesis regulators, tissue

CC growth regulators, regulators of reproductive hormones or cell movement

CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or

CC tumour inhibition activity. The DNAs can be used in forensic procedures

CC to identify individuals or in diagnostic procedures to identify

CC individuals having genetic diseases resulting from abnormal expression of

CC the genes corresponding to the extended cDNAs. They are also useful for

CC constructing a high resolution map of the human chromosomes. They can

CC also be used for gene therapy to control or treat genetic diseases.

XX SQ Sequence 121 AA;

Query Match 100.0%; Score 666; DB 20; Length 121;

Best Local Similarity 100.0%; Pred. No. 7.7e-58;

Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLPQWLLLLLFLFFFLTRGSLSPKYNLLEKESCIRNQDCETGCCORAPDNCEH 60

Db 1 mmlpqwlllllflfflltrgslsptkynlllelkesclrnqdcetgccqrapdnceh 60

QY 61 CAEKGSGSLCQTQVFFGYRACPCRLNLTCTIYSKNEKWLSIAYGRCKIGROKLAKKMF 120

Db 61 caekgsgslcqtqvfgyracpcrlnlctiysknekwlslaygrocklgrqklakkmf 120

QY 121 F 121

Db 121 f 121

RESULT 2

Y91513

ID Y91513 standard; Protein: 122 AA.

XX Y91513;

XX 29-JUN-2000 (first entry)

XX Human secreted protein sequence encoded by gene 63 SEQ ID NO:186.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

KW antiHIV; antiinflammatory; neutropic; neuroprotective; antiallergic;

KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;

KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;

KW immune disease; inflammation; blood disorder; tumour.

XX Homo sapiens.

XX WO200006698-A1.

XX 10-FEB-2000.

XX 29-JUL-1999; 99WO-US17130.

XX 30-JUL-1998; 98US-0094657.

PR 05-AUG-1998; 98US-0095486.

PR 06-AUG-1998; 98US-0095454.

PR 06-AUG-1998; 98US-0095455.

PR 12-AUG-1998; 98US-0096319.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

XX WPI; 2000-195282/17.

DR N-PSDB; A26408.

XX New isolated human genes and the secreted polypeptides they encode,

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders -

XX Claim 11; Page 499; 634pp; English.

XX The polynucleotide sequences given in A26346 to A26458 encode the human

CC secreted proteins given in Y91451 to Y91691. The human secreted proteins

CC can have activities based on the tissues and cells they are expressed in.

CC Examples of the activities are: cytostatic; immunosuppressive; antiHIV;

CC antinflammatory; neutropic; neuroprotective; antiallergic; osteopathic;

CC antiarthritic; antibacterial; antidiabetic; antiasthma; antipsoriatic;

CC and cardiant. The polynucleotides and their corresponding secreted

CC proteins are useful for preventing, treating or ameliorating medical

CC conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the proteins in a sample or

CC by determining the presence of mutations in the polynucleotides. Specific

CC uses are described for each of the polynucleotides, based on which

CC tissues they are most highly expressed in, and include developing

CC products for the diagnosis or treatment of cancer, tumours, and foetal

CC neurodegenerative disorders, developmental abnormalities and foetal

CC deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal diseases, schizophrenia,

CC allergies, Alzheimer's and behavioural disorders, inflammation,

CC osteoporosis, arthritis, infections, AIDS, sepsis, acne, psoriasis,

CC transplant rejection, diabetes, asthma, spinal cord injuries,

CC cardiovascular disorders, reproductive disorders, gastrointestinal

CC disorders, respiratory disorders and metabolic disorders. The proteins

CC or polynucleotides can also be used as food additives or preservatives.

CC The proteins are also useful for identifying their binding partners.

CC A26337 to A26345 and Y91450 are sequences used in the exemplification of

CC the present invention.

XX SQ Sequence 122 AA;

Query Match 100.0%; Score 666; DB 21; Length 122;

Best Local Similarity 100.0%; Pred. No. 7.8e-58;

Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLPQWLLLLLFLFFFLTRGSLSPKYNLLEKESCIRNQDCETGCCORAPDNCEH 60

Db 1 mmlpqwlllllflfflltrgslsptkynlllelkesclrnqdcetgccqrapdnceh 60

QY 61 CAEKGSGSLCQTQVFFGYRACPCRLNLTCTIYSKNEKWLSIAYGRCKIGROKLAKKMF 120

Db 61 caekgsgslcqtqvfgyracpcrlnlctiysknekwlslaygrocklgrqklakkmf 120

QY 121 F 121

Db 121 f 121

RESULT 3

Y12687

ID Y12687 standard; Protein: 120 AA.

XX Y12687;

XX 21-JUN-1999 (first entry)

XX Human 5' EST secreted protein SEQ ID NO:277.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

XX WO9006549-A2.

XX 11-FEB-1999.

XX 31-JUL-1998; 98WO-1B01231.

XX 01-AUG-1997; 97US-0905279.

XX (GEST) GENSET.

XX Duclert A, Dumas Milne Edwards J, Lacroix B;

XX WPI: 1999-153779/13.

XX N-PSDB; X51465.

XX New nucleic acids encoding human secreted proteins - obtained from
cDNA libraries derived from testis, ovary, uterus and spleen tissue

XX Claim 34; Page 387; 522pp; English.

XX X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
secreted proteins, and encode the proteins given in Y12681 to Y12913,
respectively. The proteins given represent the signal peptide and an
N-terminal fragment of a secreted protein. The nucleic acid sequences
can be used for producing secreted human gene products. The proteins
be used to develop products for diagnosis and therapy. They can also
obtained may have cytokine activity, cell proliferation/differentiation
activity, haematopoiesis regulating activity, tissue growth regulating
activity, reproductive hormone regulating activity, chemotactic/
chemokinetic activity, haemostatic and thrombolytic activity, receptor/
ligand activity, anti-inflammatory activity, tumour inhibition activity
or other activities. The products can be used in forensic, gene therapy
and chromosome mapping procedures. The sequences can also be used for
obtaining corresponding promoter sequences. The nucleic acids encoding
the signal peptide can be used for directing extracellular secretion of
a polypeptide or the insertion of a polypeptide into a membrane, or
importing a polypeptide into a cell.

XX Sequence 120 AA;

Query Match 92.9%; Score 619; DB 20; Length 120;

Best Local Similarity 95.8%; Pred. No. 2.8e-53;

Matches 113; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MMLPQWLLLLFFLLTRGSLSPKYNLLEKESCIRNQCETGCCQAPDNCESH 60

Db 1 mmlpqwlllllfflltrgslspkynlllelxxxxngndcctgcccqrapdncesh 60

QY 61 CAEKGSGSLCQTQVFGQYRACPLRNLTCTIYSKNEKWLSTIAYGRQKTRQKLAKK 118

Db 61 caekgsgslcqtqvffggyracplrnltctiysknekwlsiavglrcqkigrqklark 118

RESULT 4

Y44934

ID Y44934 standard; Protein; 271 AA.

XX AC Y44934;

XX 23-MAY-2000 (first entry)

XX Human dickkopf-1 homolog 3 protein.

XX Human: dickkopf-1 homolog 3; DDKh-3; Wnt antagonist; probe; primer;

XX detect; quantitate; screening assay; immunogen; antibody.

XX Homo sapiens.

FH Key Location/Qualifiers

FT Region 208..271

XX /note= "This region is claimed as SEQ ID NO: 2"

XX WO200006714-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US16963.

XX 31-JUL-1998; 98US-0094947.

XX 31-JUL-1998; 98US-0094976.

XX 31-JUL-1998; 98US-0095217.

XX (ELIL) LILLY & CO ELI.

XX Edmonds BT;

XX WPI: 2000-195288/17.

XX Novel isolated Dickkopf-1 homolog 3 nucleic acid useful as a probe or
primer for detecting, quantitating and isolating gene transcripts in
screening assays

XX Claim 6; Page 70; 72pp; English.

XX The present sequence is a human dickkopf-1 homolog 3 (DDKh-3) protein
which is selectively expressed in endocrine, and gastrointestinal tissues.
CC musculoskeletal, nervous, reproductive and gastrointestinal tissues.
CC It acts as Wnt antagonist. Dickkopf-1 has potent head inducing activity
in Xenopus. The DDKh-3 protein is used as immunogens in preparing
CC antibodies. The corresponding polynucleotide is used as a probe or primer
CC for detecting, quantitating or isolating gene transcripts e.g. in
CC screening assays for detecting allelic variants, or mutations. It is also
CC used for sense or antisense suppression of one or more DDKh-3 genes and
CC for identifying compounds that modulate transcription and translation
of the DDKh-3 genes.

XX Sequence 271 AA;

Query Match 12.8%; Score 85.5; DB 21; Length 271;

Best Local Similarity 26.1%; Pred. No. 0.46;

Matches 23; Conservative 9; Mismatches 29; Indels 27; Gaps 5;

QY 39 CIRNQDCEGCGCQAPDNCESH-----CAEKGSGSLCQTQVFGQ-----YRACPLRN 88

Db 195 clrssdcaaglc-----carhfwtkickpvlhgdgevtckqrkkgshgleifgrcdcaqg 248

QY 89 LTCIYSKNEKWLSTIAYGR-----CQKI 110

Db 249 lsc-----kvwkdatyskharlhvcqki 271

RESULT 5

W73019

ID W73019 standard; Protein; 179 AA.

XX AC W73019;

XX 18-JAN-1999 (first entry)

XX Human cysteine-rich secreted protein CRSP-4.

XX CRSP-4; cysteine-rich secreted protein 4; tumour; cancer; leukaemia;

XX tissue repair; wound healing; infection; Parkinson's disease;

XX Alzheimer's disease; Huntington's chorea; multiple sclerosis;

XX amyotrophic lateral sclerosis; pontine myelinolysis;

XX human immunodeficiency associated myelopathy; bulbar palsy;

XX spinal muscular atrophy; primary lateral sclerosis; poliomyelitis;

XX Razio-Londe syndrome; Charcot-Marie-Tooth disease; therapy;

XX diagnosis; drug screening; human; signal transduction;

XX cell differentiation; cell proliferation.

Query Match 12.4%; Score 82.5; DB 21; Length 207;
Best Local Similarity 25.6%; Pred. No. 0.69;
Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
QY 37 ESIRNODCTGCGQAPDNCEH-----CAEKSGESLCQTVFFGQ-----YRACPCL 86
Db 129 dpcIrsdcIegfc-----carhfwtkickpvhgvevctkqkkgshgleifqrcdca 182
QY 87 RNLTCIYKNEKWLSTAYGR-----CQKI 110
Db 183 kgIsc-----kvwkdatyskcarlhvcqkl 207

RESULT 7

Y99360 ID Y99360 standard; Protein; 259 AA.

XX AC Y99360;

XX XX

DT DT 08-AUG-2000 (first entry)

DE Human PRO1316 (UNQ682) amino acid sequence SEQ ID NO:70.

XX KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;

KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX OS Homo sapiens.

XX XX

PN PN W0200012708-A2.

XX XX

PD PD 09-MAR-2000.

PF PF 01-SEP-1999; 99WO-US20111.

XX XX

XX 01-SEP-1998; 98US-0098716.

PR PR 01-SEP-1998; 98US-0098749.

PR PR 01-SEP-1998; 98US-0098750.

PR PR 02-SEP-1998; 98US-0098803.

PR PR 02-SEP-1998; 98US-0098821.

PR PR 02-SEP-1998; 98US-0098843.

PR PR 09-SEP-1998; 98US-0099536.

PR PR 09-SEP-1998; 98US-0099596.

PR PR 09-SEP-1998; 98US-0099602.

PR PR 09-SEP-1998; 98US-0099642.

PR PR 10-SEP-1998; 98US-0099741.

PR PR 10-SEP-1998; 98US-0099754.

PR PR 10-SEP-1998; 98US-0099763.

PR PR 10-SEP-1998; 98US-0099792.

PR PR 10-SEP-1998; 98US-0099808.

PR PR 10-SEP-1998; 98US-0099812.

PR PR 10-SEP-1998; 98US-0099815.

PR PR 15-SEP-1998; 98US-0099816.

PR PR 15-SEP-1998; 98US-0100385.

PR PR 15-SEP-1998; 98US-0100388.

PR PR 15-SEP-1998; 98US-0100390.

PR PR 16-SEP-1998; 98US-0100584.

PR PR 16-SEP-1998; 98US-0100627.

PR PR 16-SEP-1998; 98US-0100661.

PR PR 16-SEP-1998; 98US-0100662.

PR PR 16-SEP-1998; 98US-0100664.

PR PR 17-SEP-1998; 98US-0100683.

PR PR 17-SEP-1998; 98US-0100684.

PR PR 17-SEP-1998; 98US-0100710.

PR PR 17-SEP-1998; 98US-0100711.

PR PR 17-SEP-1998; 98US-0100919.

PR PR 17-SEP-1998; 98US-0100930.

PR PR 18-SEP-1998; 98US-0100848.

PR PR 18-SEP-1998; 98US-0100849.

PR PR 18-SEP-1998; 98US-0101014.

PR PR 18-SEP-1998; 98US-0101068.

PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101913.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.

XX PD 06-APR-2000.
 XX PF 17-SEP-1999; 99WO-US21647.
 XX PR 25-SEP-1998; 98US-0161241.
 XX PA (AMGE-) AMGEN INC.
 XX PI Bass MB, Sullivan JK, Theill LE, Wang D;
 XX PT WPI; 2000-293153/25.
 XX DR N-PSDB; A08842.
 XX PT New nucleic acid molecule encoding a biologically active DKR
 XX PT polypeptide, useful in treatment of cancer, e.g. mammary tumors and
 XX PT stem cell tumors
 XX PS Claim 18; Page 129-130; 143pp; English.
 XX CC Y92069-75 are novel mouse and human DKR polypeptides.
 CC The human DKR-2 gene has about 95 percent homology to murine DKR-2.
 CC DKR-1 is a human ortholog of dkk-1 (dickkopf-1), a novel gene identified
 CC in Xenopus and mouse, purportedly an antagonist of wnt-8 signaling.
 CC DKR-2, -3 and -4 are each related to DKR-1 by their cysteine pattern.
 CC Dkk-1 is also involved in morphogenesis in the developing embryo, and
 CC therefore a growth factor, by inference DKR polypeptides are also
 CC growth factors. The DKR polypeptides are useful for treating cancer,
 CC e.g. mammary tumors, stem cell tumors, or other cancers in which the wnt
 CC and/or sonic hedgehog (shh) signal transduction pathways are activated.
 CC They can also be used to enhance tissue differentiation, such as bone
 CC formation and hematopoietic cell formation.
 XX CC
 XX SQ Sequence 259 AA;
 Query Match 12.4%; Score 82.5; DB 21; Length 259;
 Best Local Similarity 25.6%; Pred. No. 0.87;
 Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;
 QY 37 ESCIRNDCETGCCORAPDNCEH-----CAEKGSEGLCQTQVFFGQ-----YRACPCL 86
 Db 181 dpcrlrsddclegfc-----carhfwtkickpvlhggevtckqrkkgshgleifqrdca 234
 QY 87 RNLTCTIYSKNEKWLSTIAYGR-----CQKI 110
 Db 235 kgisc-----kvwkdatysksarlhvcqki 259
 RESULT 10
 R25449
 ID R25449 standard; Protein; 125 AA.
 XX AC R25449;
 XX DT 13-JAN-1993 (first entry)
 XX DE PDGF analogue #9.
 XX KW platelet derived growth factor; chemotactic; mitogenic; fibroblasts;
 KW wound healing; dermal ulcers; lacerations; abrasions;
 KW surgical wounds; burns; defined culture media; v-sis protein; p28sis;
 KW neoplasm; cancer; tumour; inhibit atherosclerosis.
 XX OS Synthetic.
 XX PN US5128321-A.
 XX PD 07-JUL-1992.
 XX PF 08-AUG-1988; 88US-0230190.
 XX PR 12-OCT-1984; 84US-0660496.

PR 25-FEB-1985; 85US-0705175.
 PR 13-AUG-1986; 86US-0896485.
 PR 15-DEC-1986; 86US-0941970.
 XX (ZYMO) ZYMOGENETICS INC.
 XX PI Kelly JD, Murray MJ;
 XX DR WPI; 1992-249442/30.
 XX PT New platelet-derived growth factor analogues - comprising protein
 PT homo:dimers having two polypeptide chains with sequences the same
 PT as the A- and opt. B chain obtd. in large amts. used to enhance the
 PT healing of wounds
 XX PS Disclosure; Fig 3A; 25pp; English.
 XX CC This sequence represents a new platelet-derived growth factor
 CC analogue chain . The analogue functions as a homodimer of two chains,
 CC each of the chains being a mosaic of amino acid sequences identical
 CC to portions of the A- or B-chains of PDGF, the protein being
 CC chemotactic or mitogenic for fibroblasts. The homodimers have the
 CC same biological activity as PDGF, but can be obtained in large
 CC amts. by recombinant DNA techniques. These PDGF analogues can be
 CC used in a compsn. to enhance the healing of wounds eg. dermal
 CC ulcers, superficial wounds and lacerations, abrasions, surgical
 CC wounds and some burns. The PDGF analogues can also be used as
 CC components of defined, serum-free culture media. They can also be
 CC used in elucidating the putative role of the vsis protein p28sis
 CC in the neoplastic process or for developing inhibitors or
 CC designing specific therapeutic approaches which prevent or
 CC interfere with the in vivo activity of PDGF in individuals with
 CC atherosclerosis.
 XX CC
 XX SQ Sequence 125 AA;
 Query Match 11.6%; Score 77; DB 13; Length 125;
 Best Local Similarity 22.8%; Pred. No. 1.4;
 Matches 26; Conservative 22; Mismatches 48; Indels 18; Gaps 5;
 QY 17 LFLTRGSLSPKYNLLEKESCIRNDCETGCCORAPDNCE-----SHCAEK-----G 65
 Db 16 iyeiprsqvdptsanfl-iwppcvevkrc-tgcqctssvkcpqsrhhrsvkvakveyvr 73
 QY 66 SEGSLCQTQVFFGQYRACPCLRNLTCTIYSKNEKWLSTIAYGECQKIGROKLAKKM 119
 Db 74 kpkkiqevqrleehleacatt-----slndpyreedtgrpresgkkrkrkl 122
 RESULT 11
 R87521
 ID R87521 standard; peptide; 125 AA.
 XX AC R87521;
 XX DT 21-JUN-1996 (first entry)
 XX DE Glioma cell PDGF A chain.
 XX KW Platelet-derived growth factor; PDGF; unglycosylated protein; burn;
 KW wound healing; non-healing dermal ulcer; superficial wound; laceration;
 KW abrasion; surgical wound; advanced age; diabetes; cancer; steroid;
 KW anti-inflammatory drug; anticoagulant; homodimer.
 XX OS Synthetic.
 XX PN US5474982-A.
 XX PD 12-DEC-1995.
 XX PF 13-AUG-1986; 86US-0896485.
 XX XX

PR 08-AUG-1988; 88US-0230190.
 PR 13-AUG-1986; 86US-0896485.
 PR 15-DEC-1986; 86US-0941970.
 PR 30-JUN-1992; 92US-0906544.
 PR 12-JUL-1994; 94US-0273779.
 XX (ZYMO) ZYMOGENETICS INC.
 XX Kelly JD, Murray MJ;
 PI WPI; 1996-039529/04.
 XX New platelet-derived growth factor analogues - useful for promoting
 PT wound healing
 XX Example 3; Fig 3b; 25pp; English.
 PS R87513-R87521 are representative peptides of the invention. These
 CC sequences are all mosaics of the platelet-derived growth factor (PDGF) A
 CC chain. This sequence represents a glioma cell A chain variant, which can
 CC be used as a peptide of the invention. These sequences are used to form
 CC the novel unglycosylated protein homodimers of the invention. These
 CC homodimers can be used in compositions for enhancing the wound healing
 CC process. The protein is expected to accelerate the healing process in a
 CC broad spectrum of wound conditions including disruptions of the dermal
 CC layer of the skin (such as non-healing dermal ulcers, superficial wounds
 CC and lacerations, abrasions, surgical wounds, and some burns). These
 CC sequences are especially used in enhancing the wound-healing process in
 CC conditions where the normal wound healing process is suppressed or
 CC inhibited, such as advanced age, diabetes, cancer, and treatment with
 CC anti-inflammatory drugs, steroids or anticoagulants.
 XX Sequence 125 AA;

Query Match 11.6%; Score 77; DB 17; Length 125;
 Best Local Similarity 22.8%; Pred. No. 1.4;
 Matches 26; Conservative 22; Mismatches 48; Indels 18; Gaps 5;
 QY 17 LFLTRGSLPTKYNLLEKESCIRNODCETGCCQAPDNCE----SHCAEK-----G 65
 DB 16 iyeiprsqvdtsanfl-iwpcvevkrc-tgcqtsvkcqpsrvhrsvkvakveyvr 73
 QY 66 SEGSLCQTVFGQVRACPLRLNLTCTIYSKNEKWLSTAYGRCQKIGROKLAKKM 119
 DB 74 kfpklkevqvrieehleacatt-----slpdyreedtgrpresgkkrkrkrl 122

RESULT 12
 R44242
 ID R44242 standard; Protein; 802 AA.
 AC R44242;
 DT 03-MAY-1994 (first entry)
 XX F-spondin of chicken.
 DE Thrombospondin; TSR; Thrombospondin type I repeat; F-spondin;
 KW neurodevelopment; nerve; axon; adhesion; outgrowth.
 KW Gallus gallus.
 OS WO9320196-A.
 PN 14-OCT-1993.
 XX 02-APR-1993; 93WO-US03164.
 XX 02-APR-1992; 92US-0862021.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX

PI Jessell TM, Klar A;
 XX WPI; 1993-336904/42.
 DR N-PSDB; Q52675.
 XX New vertebrate F-spondin protein - used for attaching nerve cells
 PT to a matrix, stimulating growth of nerve cells or regenerating
 PT nerve cells
 XX Claim 18; Page 68-70; 103pp; English.
 XX F-spondin is useful for adhesion and outgrowth of axons. It can be
 CC used for attaching nerve cells to a matrix, stimulating growth of
 CC nerve cells or regenerating nerve cells. F-spondin nucleic acid
 CC can be used to develop probes to study neurodevelopment.
 CC Antibodies to the F-spondin can be used for determining the
 CC localisation of the protein in the nervous system and in assessing
 CC its function.
 XX Sequence 802 AA;
 SQ Query Match 11.4%; Score 76; DB 14; Length 802;
 Best Local Similarity 21.8%; Pred. No. 12;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;
 QY 34 ELKESCIRNODCETGCCQAP----DNCESHCAE-----KGSEGLCQTVFGF 78
 DB 538 eetekclvneecspssclvtewgedscascgtgmkrhrmiktmgadgsnckaeetee 597
 QY 79 QYRACPLRLNLTCTIYSKNEKWLSTAYGRCQKIGROKLAK 117
 DB 598 ekcmpechtipcllspwsewscsvtgcgkmrt-rgrmlk 637

RESULT 13
 R56166
 ID R56166 standard; Protein; 383 AA.
 XX R56166;
 DT 26-JAN-1995 (first entry)
 XX Neuroendocrine tumor dlk.
 DE Dlk; neuroendocrine tumor; monoclonal antibody.
 KW Homo sapiens.
 OS Key Location/Qualifiers
 FH Peptide 1..23
 FT /label= Sig_peptide
 FT Modified-site 62
 FT /label= N-myristylation_site
 FT Modified-site 72
 FT /label= N-myristylation_site
 FT Modified-site 100
 FT /label= N-glycosylation_site
 FT Modified-site 101
 FT /label= N-myristylation_site
 FT Modified-site 109
 FT /label= N-myristylation_site
 FT Modified-site 120
 FT /label= protein-kinase-C-phosphorylation_site
 FT Modified-site 187
 FT /label= hydroxylation_site
 FT Modified-site 288
 FT /label= protein-kinase-C-phosphorylation_site
 FT Modified-site 302
 FT /label= N-myristylation_site
 FT Modified-site 312
 FT /label= N-myristylation_site
 XX

PN WO9413701-A.

XX 23-JUN-1994.

XX 10-DEC-1993; 93WO-US12015.

XX 11-DEC-1992; 92US-0989537.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX (USSH) US SEC DEPT. HEALTH.

XX Laborda J;

XX WPI; 1994-217806/26.

XX N-PSDB; Q68221.

XX A Dlk polypeptide encoded by dlk polynucleotide mol - can be used

PT in detecting the existence of primary or secondary neuro

PT endocrine tumour

XX Disclosure; Fig. 1; 24pp; English.

XX A cDNA clone was isolated from a phage lambda-gtl0 human adrenal

CC gland library using mouse dlk as hybridization probe. The human

CC (Q68221) and mouse (Q68222) dlk cDNAs showed 82.2% identity. The

CC encoded human (R56166) and mouse (R56167) Dlk proteins are used to

CC detect neuroendocrine tumors and to produce monoclonal antibodies

CC useful for diagnosis and therapy.

XX Sequence 383 AA;

XX Query Match 11.3%; Score 75; DB 15; Length 383;

XX Best Local Similarity 34.4%; Pred. No. 6.9;

XX Matches 21; Conservative 6; Mismatches 22; Indels 12; Gaps 5;

QY 35 LKESCIRNQDCETGCCORAPDNCSHCARSGSLCOTQVFFQYRAC---PCLRNLTG 91

DB 53 lcdqcvspgclhgcge-p9qc--ictd-gwdgelcdrdv-----racssapcanngtc 103

QY 92 I 92

DB 104 v 104

RESULT 14

W87895

ID W87895 standard; Protein; 1148 AA.

XX W87895;

XX 26-APR-1999 (first entry)

XX Human JAGGED2 protein.

XX JAGGED2; JAGGED2; human; notch ligand; syndactyl;

XX cleaft palate; cleft lip; diagnosis; stem cell;

XX progenitor cell; haematopoiesis; cell differentiation.

XX Homo sapiens.

XX WO9858958-A2.

XX 30-DEC-1998.

XX 25-JUN-1998; 98WO-US13207.

XX 25-JUN-1997; 97US-0882046.

XX (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.

XX (UNIW) UNIV WASHINGTON.

XX Hood L, Krantz ID, Li L, Spinner NB;

PI

XX

DR WPI; 1999-081220/07.

XX N-PSDB; V63754.

XX New Jagged peptides for inhibiting differentiation of progenitor

PT cells - also used for maintaining these cells in undifferentiated

PT state, e.g. for haematopoietic reconstitution

XX Claim 6; Fig 1B; 101pp; English.

XX This is the amino acid sequence of human JAGGED2 (hJAGGED2), a

CC membrane-bound ligand with a large extracellular domain and a

CC relative short intracellular domain. The polypeptide also has a

CC Delta/Serrate/Lag-2 domain, 15 EGF-like repeats and a

CC transmembrane domain characteristic of membrane-bound Notch

CC ligands. Genetic polymorphisms in the human JAGGED2 gene (see

CC V63754) are associated with syndactyl and cleft palate or lip.

CC A method of diagnosing a syndrome characterised by such conditions

CC is disclosed. The invention also relates to new human JAGGED2

CC protein (see W87894) and polynucleotides (see V63753) and their

CC used for inhibiting progenitor cell differentiation and diagnosing

CC Alagille syndrome.

XX Sequence 1148 AA;

XX Query Match 11.3%; Score 75; DB 20; Length 1148;

XX Best Local Similarity 23.2%; Pred. No. 21;

XX Matches 23; Conservative 11; Mismatches 47; Indels 18; Gaps 4;

QY 31 NLELEKESCIRNQDCETGCCORAPDNCSHCARSGSLCOTQVFFQYRACPLRNLT 90

DB 293 nckaeahactsnpcanggschevpsgfechc-psgwsqptcaldi--decasnpscaagt 349

QY 91 CIYSKN-----EKWLSI-----AYRCQKIGROK 114

DB 350 cvdqvdfecicpeqwgatcqldvndrgcgqbggtck 388

RESULT 15

R56167

ID R56167 standard; Protein; 385 AA.

XX R56167;

XX 26-JAN-1995 (first entry)

XX Neuroendocrine tumor dlk.

XX Dlk; neuroendocrine tumor; monoclonal antibody.

XX Mus sp.

XX Key

XX Peptide

XX Location/Qualifiers

XX 1..23

XX /label= Sig_peptide

XX Modified-site

XX 62

XX /label= N-myristylation_site

XX Modified-site

XX 100

XX /label= N-glycosylation_site

XX Modified-site

XX 109

XX /label= N-myristylation_site

XX Modified-site

XX 120

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 189

XX /label= hydroxylation_site

XX Modified-site

XX 269

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 290

XX /label= protein-kinase-C_phosphorylation_site

XX Modified-site

XX 295

XX /label= N-glycosylation_site

XX Modified-site

XX 304


```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Monaghan P.A., Kioschis P., Wu W., Zuniga A., Bock D., Poustka A.,
RA Delius H., Niehrs C.;
RT "Dickkopf genes are co-ordinately expressed in mesodermal lineages.";
RL Mech. Dev. 87:45-56(1999).
DR EMBL: AJ243963; CAB60110.1; -.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 259 DICKKOPF-2.
SQ SEQUENCE 259 AA; 28416 MW; EAAB76F2D2C9780D CRC64;

Query Match 12.4%; Score 82.5; DB 11; Length 259;
Best Local Similarity 25.6%; Pred. No. 0.064;
Matches 23; Conservative 10; Mismatches 30; Indels 27; Gaps 5;

QY 37 ESCIRNDCETGCCORAPDNCEH-----CAEKSGEGLCQTQVFFGQ-----YRACPCL 86
Db 181 DPLRSDDICDGF-----CARHFWTKICKPVLHOGVCTKORKKGSHGLEIFQRCDCA 234

QY 87 RNLTCTYSKNEKWLSTAYGR-----CQKI 110
Db 235 KGLSC-----KVKWDATYSSKARLHVCQKI 259

RESULT 6
ID 070534 PRELIMINARY; PRT; 383 AA.
AC 070534;
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
DE ZOG.
GN ZOG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Halder S.K., Takemori H., Akira W., Yasuki N., Mitsuhiro O.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; D84336; BAA25881.1; -.
DR HSSP; P00740; 11XA.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF_5.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS01186; EGF_2; 6.
KW Glycoprotein.
SQ SEQUENCE 383 AA; 41149 MW; 123724C9801A9FD5 CRC64;

Query Match 12.3%; Score 82; DB 11; Length 383;
Best Local Similarity 31.1%; Pred. No. 0.11;
Matches 19; Conservative 9; Mismatches 21; Indels 12; Gaps 4;

QY 35 LKESCIRNDCETGCCORAPDNCEHCAEKSGEGLCQTQVFFGQYRAC---PCLRNLTG 91
Db 53 LCEKCVTSPGCVNGLCEE-PWQC---VCKEGWDGKFCFCEIDI-----RACTSTPCANNGTC 103

QY 92 I 92
Db 104 V 104

RESULT 8
QYHU2
ID QYHU2 PRELIMINARY; PRT; 1254 AA.
AC QYHU2;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE SERRATEB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Haddon C., Jiang Y.-J., Smithers L., Lewis J.;
RT "Delta-Notch signalling and the patterning of sensory cell
```

```
RESULT 7
Q62779
ID Q62779 PRELIMINARY; PRT; 383 AA.
AC Q62779;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
DE PREADIPOCYTE FACTOR 1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=WISTAR; TISSUE-ISLETS OF LANGERHANS;
RX MEDLINE-97419009; PubMed-9275085;
RA Carlsson C., Tornehave D., Lindberg K., Galante P., Billestrup N.,
RA Michelsen B., Larsson L.I., Nielsen J.H.;
RT "Growth hormone and prolactin stimulate the expression of rat
RT preadipocyte factor-1/delta-like protein in pancreatic islets:
RT molecular cloning and expression pattern during development and growth
RT of the endocrine pancreas.";
RL Endocrinology 138:3940-3948(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=WISTAR; TISSUE-ISLETS OF LANGERHANS;
RA Svensson C.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25680; AAB87095.1; -.
DR HSSP; P00740; 11XA.
DR INTERPRO; IPR000152; -.
DR INTERPRO; IPR000561; -.
DR INTERPRO; IPR000742; -.
DR INTERPRO; IPR001881; -.
DR PFAM; PF00008; EGF_5.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_5.
DR PROSITE; PS01186; EGF_2; 6.
KW Glycoprotein.
SQ SEQUENCE 383 AA; 41253 MW; 7A6F90C02995C100 CRC64;

Query Match 12.3%; Score 82; DB 11; Length 383;
Best Local Similarity 31.1%; Pred. No. 0.11;
Matches 19; Conservative 9; Mismatches 21; Indels 12; Gaps 4;

QY 35 LKESCIRNDCETGCCORAPDNCEHCAEKSGEGLCQTQVFFGQYRAC---PCLRNLTG 91
Db 53 LCEKCVTSPGCVNGLCEE-PWQC---VCKEGWDGKFCFCEIDI-----RACTSTPCANNGTC 103

QY 92 I 92
Db 104 V 104

RESULT 8
QYHU2
ID QYHU2 PRELIMINARY; PRT; 1254 AA.
AC QYHU2;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE SERRATEB.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi.
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Haddon C., Jiang Y.-J., Smithers L., Lewis J.;
RT "Delta-Notch signalling and the patterning of sensory cell
```

RT differentiation in the zebrafish ear: evidence from the mind bomb
 mutant.;
 RL Development 125:4637-4644(1998).
 DR EMBL; AF090432; AAC98354.1; -;
 DR HSSP; P00740; 1EDM.
 DR INTERPRO; IPR000152; -;
 DR INTERPRO; IPR000561; -;
 DR INTERPRO; IPR000742; -;
 DR INTERPRO; IPR000950; -;
 DR INTERPRO; IPR001438; -;
 DR INTERPRO; IPR001774; -;
 DR INTERPRO; IPR001881; -;
 DR PFAM; PF00008; EGF; 15.
 DR PFAM; PF01414; DSL; 1.
 DR PRINTS; P000010; EGFLOOD.
 DR PROSITE; PS00010; ASX-HYDROXYL; UNKNOWN_7.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_16.
 DR PROSITE; PS00241; RECEPTOR_CYTOKINES_1; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 6.
 KW Glycoprotein; EGF-like domain.
 SQ SEQUENCE 1254 AA; 137685 MW; BB902C35D3502B70 CRC64;

Query Match 12.2%; Score 81; DB 13; Length 1254;
 Best Local Similarity 28.4%; Pred. No. 0.41;
 Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 4;
 QY 27 PTYNL-----LEKE-SCIRNQCETCCORAPDNCESHCAEKGSGSLCQTV 75
 Db 320 PDEVNCAPEGYSGKNCIEAHACVSNPCANGGTCEVPTGFCHC-PPGWEGPTCAKDM 378
 QY 76 FFGQYRACPLRLNLTCTYSKN 96
 Db 379 --DECASSPCAQGGTCIDLEN 397

RESULT 9
 ID Q9QXU5 PRELIMINARY; PRT; 107 AA.
 AC Q9QXU5;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE BV8 VARIANT 3 PRECURSOR.
 GN BV8.
 OS Mus. musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC STRAIN=129/SV; TISSUE=TESTIS;
 RA Wechselsberger C., Puglisi R., Lepperdinger G., Boitani C., Kreil G.;
 RT "The mammalian homologue of Bv8 from frog skin is mainly expressed in
 spermatoocytes.";
 RL FEBS Lett. 462:177-181(1999).
 DR EMBL; AF182066; AAF15261.1; -;
 DR HSSP; P25687; 11MT.
 KW Signal.
 FT SIGNAL
 SQ SEQUENCE 107 AA; 11556 MW; 8FB79BE4D4357C70 CRC64;

Query Match 12.1%; Score 80.5; DB 11; Length 107;
 Best Local Similarity 25.08%; Pred. No. 0.049;
 Matches 26; Conservative 16; Mismatches 47; Indels 15; Gaps 4;
 QY 8 LLLFLFFFLFLLRGSLSPTKYNLLEKESCIRNQCETGCCORAPDNCS--HCAEKG 65
 Db 9 LLLLLLLPLLF-----TPAGDAAVITGACDKDSQCGGMCACVSIWVKSIRICPPMG 61
 QY 66 SEGSLCQTV----FFQG--YRACPLRLNLTCTYSKNEKWLISIA 103

Db 62 QVGDSCHPLTRKVPFWGRMHHTCPCLPGLACLRTSFNRTCLA 105
 RESULT 10
 ID O45448 PRELIMINARY; PRT; 126 AA.
 AC O45448; O45448;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE F35C5.12 PROTEIN.
 GN F35C5.12.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Jier M., Johnston L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister L., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.

Query Match 12.1%; Score 80.5; DB 5; Length 126;
 Best Local Similarity 23.0%; Pred. No. 0.057;
 Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;
 QY 9 LLLFLFFFLFLLRGSLSPTKYNLLEKESC--IRNQCETGC----- 49
 Db 1 MLFL--FFAFLL--SLQP-----FVRGACTFLDQACDEVCKTDFSEWYHCNGWDGFN 50
 QY 50 -----CORAPDNCSHCAEKGSEGLC--QTQVFGQYRACPLRLNLTCTC 91
 Db 51 FSCCKVEYITPLDGSICETROMACSEKCKDQSGEGFCFPLDLSHKSRLTACECFKQV 110
 QY 92 IYSKNEKWLISIAVGR 106
 Db 111 LRRKRSIQKRSYKR 125

Query Match 12.1%; Score 80.5; DB 5; Length 126;
 Best Local Similarity 23.0%; Pred. No. 0.057;
 Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;
 QY 9 LLLFLFFFLFLLRGSLSPTKYNLLEKESC--IRNQCETGC----- 49
 Db 1 MLFL--FFAFLL--SLQP-----FVRGACTFLDQACDEVCKTDFSEWYHCNGWDGFN 50
 QY 50 -----CORAPDNCSHCAEKGSEGLC--QTQVFGQYRACPLRLNLTCTC 91
 Db 51 FSCCKVEYITPLDGSICETROMACSEKCKDQSGEGFCFPLDLSHKSRLTACECFKQV 110
 QY 92 IYSKNEKWLISIAVGR 106
 Db 111 LRRKRSIQKRSYKR 125

RESULT 11
 ID Q9U0E2 PRELIMINARY; PRT; 585 AA.
 AC Q9U0E2;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
 DE SIGNAL RECEPTOR PROTEIN (FRAGMENT).

GN NOTCH.
 OS Tribolium castaneum (Red flour beetle).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
 OC Cucujiformia; Tenebrionidae; Tribolium.
 OX NCBI_TaxID=7070;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Tautz D., Lardelli M., Westin J., Tamme R.;
 RA "Embryonic expression of Tribolium Notch.";
 RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AJ005083; CAB65469.1; -;
 DR HSP; P00740; IEDM.
 DR INTERPRO: IPR000083; -;
 DR INTERPRO: IPR000152; -;
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR000742; -;
 DR INTERPRO: IPR000800; -;
 DR INTERPRO: IPR001010; -;
 DR INTERPRO: IPR001438; -;
 DR INTERPRO: IPR001881; -;
 DR INTERPRO: IPR002049; -;
 DR PFAM; PF00008; EGF_11.
 DR PFAM; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFBLAMININ.
 DR PRINTS; PR00012; FNTYPEI.
 DR PRINTS; PR00287; THIONIN.
 DR PROSITE; PS00010; ASX_HYDROXYL; 6.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_12.
 DR PROSITE; PS01186; EGF_2; 9.
 DR PROSITE; PS01187; EGF_CA; 4.
 KW Receptor.
 FT NON_TER 1
 FT NON_TER 585
 FT SEQUENCE 585 AA; 63337 MW; 4CF7A51D0820D048 CRC64;

Query Match 12.0%; Score 80; DB 5; Length 585;
 Best Local Similarity 25.8%; Pred. No. 0.27;
 Matches 17; Conservative 16; Mismatches 25; Indels 8; Gaps 3;
 QY 32 LLELKESICR-----NDCEGTCCORAPDNCESHCAGKSGSLCQTVFFGQYRACPC 86
 Db 138 MVSKDAALRKXVPLKLLCNNGTCEDIGNSHRCHSD-GYSGSYCQFEI--NECDAPCQ 194
 QY 87 RNLFCI 92
 Db 195 NGGTCL 200

RESULT 12
 ID Q9XWD6 PRELIMINARY; PRT; 1111 AA.
 AC Q9XWD6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DE Y47H9C.4 PROTEIN.
 DE Y47H9C.4
 GN Caenorhabditis elegans.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94150718; PubMed=7906398;
 RA Harris B.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Saldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL; AL032657; CAA21739.1; -;
 DR HSP; P10968; TWGA.
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR002049; -;
 DR PFAM; PF00008; EGF_11.
 DR PRINTS; PR00011; EGFAMININ.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE; PS01186; EGF_2; 11.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F374C008F9874 CRC64;

Query Match 12.0%; Score 80; DB 5; Length 1111;
 Best Local Similarity 31.2%; Pred. No. 0.48;
 Matches 24; Conservative 8; Mismatches 27; Indels 18; Gaps 6;
 QY 44 DCETGC-CORAP--DNCESHC-AEKGSGSLCQTV---FFGQ--YRACPCRLNLTCTIYS 94
 Db 244 ECKFECCNGCATCDNTNGKICKSGYHGALCENECVFFGSGCTKCCLNQNCDSS 303
 QY 95 KNEKWLSTAYGRCKIG 111
 Db 304 S-----GECKCIG 311

RESULT 13
 Q28625 PRELIMINARY; PRT; 474 AA.
 AC Q28625;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE TUBULOINTERSTITIAL NEPHRITIS ANTIGEN.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=KIDNEY CORTEX;
 RC MEDLINE=95332335; PubMed=7608193;
 RA Nelson T.R., Charonis A.S., McIvor R.S., Butkowski R.J.;
 RA "Identification of a cDNA encoding tubulointerstitial nephritis
 RT antigen.";
 RL J. Biol. Chem. 270:16265-16270(1995).
 DR EMBL; U24270; AAC48477.1; -;
 DR HSP; P00787; 1THE.
 DR INTERPRO: IPR000169; -;
 DR INTERPRO: IPR000668; -;
 DR INTERPRO: IPR001212; -;
 DR PFAM; PF00112; Peptidase_C1; 2.
 DR PFAM; PF01033; Somatomedin_B; 1.
 DR PRINTS; PR00705; PAPAIN.
 DR PROSITE; PS00640; THIOI_PROTEASE_ASN; 1.
 SQ SEQUENCE 474 AA; 54521 MW; 46AFA48FB7C29D1B CRC64;

Query Match 11.6%; Score 77.5; DB 6; Length 474;
 Best Local Similarity 25.0%; Pred. No. 0.43;
 Matches 28; Conservative 18; Mismatches 43; Indels 23; Gaps 7;

CC -1- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF 6-
CC 20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 KDA).
CC EMBL: X63338; CAA44938.1; -.
DR INTERPRO: IPR002494; -.
DR PFAM: PF01500; Keratin_B2; 1.
DR Keratin.
SQ SEQUENCE 175 AA; 18218 MW; 0CB8BEFC9704837E CRC64;

Query Match 11.6%; Score 77; DB 4; Length 175;
Best Local Similarity 44.4%; Pred. No. 0.2;
Matches 16; Conservative 3; Mismatches 11; Indels 6; Gaps 2;
QY 38 SCIRNQCETGCCQRPDNCESHCAEKSGSLCQT 73
DB 23 SCCQSPCCETSCCQ--PSCCETSCCQ-----SCCQT 52

Search completed: March 24, 2001, 13:25:43
Job time: 249 sec

QY 6 WLLLLFFLLFFELP-----LLTRGSLSPTKYNLE---LKESCIRNQCETGCC 50
DB 2 WTGYKFFIYLAKEIWTQKQKQVDLASYSRNHSILEGTRFKRSIFQGYCRSLGCC 61
QY 51 QRAPDNCESHCAEKSGSLCQTQVFFGOYRA--CPCLRNLCIYSKNEKWL 100
DB 62 EGRNDCVCTQFYE---ANALCYCKFCERENSDCCPDYKSF-C--QEEKEWL 107

RESULT 14
Q49549 PRELIMINARY; PRT; 1187 AA.
AC Q49549;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE REPEAT REGIONS IN POTENTIAL METAL BINDING PROTEIN GENE REGION.
OS Mycoplasma hyorhinis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GBL1;
RX MEDLINE=95014025; PubMed=7928953;
RA Deng G., McIntosh M.A.;
RT "An amplifiable DNA region from the Mycoplasma hyorhinis genome."
RL J. Bacteriol. 176:5929-5937(1994).
DR EMBL: L11447; AAA62228.1; -.
SQ SEQUENCE 1187 AA; 135027 MW; 66A82AE5B0EDE93E CRC64;

Query Match 11.6%; Score 77.5; DB 2; Length 1187;
Best Local Similarity 36.2%; Pred. No. 1;
Matches 25; Conservative 7; Mismatches 26; Indels 11; Gaps 6;
QY 27 PTKYNLLKESCIRNQCETGCCQRPDNCESHCAEK--GSEGLCQ-TQVFFG-QYRA 82
DB 339 PGCSLKETEESC-----DCEACKQCEECNC--CSELTGCGQCATSCAQHCGCQDES 392
QY 83 CPCLRNLC 91
DB 393 CAC-PNTTC 400

RESULT 15
Q07628 PRELIMINARY; PRT; 175 AA.
AC Q07628;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE KERATIN.
GN HB2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93062886; PubMed=1279380;
RA Zhunabaeva B.D., Gening L.V., Gazaryan K.G.;
RT "[Cloning and structural characteristics of human hair keratin genes
rich in sulfur].";
RL Mol. Biol. (Mosk) 26:550-555(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93062886; PubMed=1279380;
RA Zhunabaeva B.D., Gening L.V., Gazaryan K.G.;
RT "[Cloning and structural characteristics of human hair keratin genes
rich in sulfur].";
RL Mol. Biol. (Mosk) 26:813-820(1992).

THIS PAGE BLANK (USPTO)

A:Accession: A42568
A:Molecule type: DNA
A:Residues: 1-112 <SIM>
A:Cross-references: GB:M95529; NID:q180842; PIDN:AA05818.1; PID:q1483624
A:Note: Sequence extracted from NCBI backbone (NCBIN:110576, NCBIN:110580)
R:Low, M.E.; Rosenblum, J.L.; McEwen, P.; Strauss, A.W.
Biochemistry 29, 823-828, 1990
A:Title: Cloning and characterization of the human colipase cDNA.
A:Reference number: A33949; MUID:90248429
A:Accession: A33949
A:Molecule type: mRNA
A:Residues: 1-112 <LOW>
A:Cross-references: GB:J02883; NID:q180885; PIDN:AAA52054.1; PID:g180886
A:Note: Evidence of partial N-glycosylation, possibly at Asn-43
R:Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A:Title: The primary sequence of human pancreatic colipase.
A:Reference number: A90652; MUID:84104937
A:Accession: A03163
A:Molecule type: protein
A:Residues: 23-108 <STE>
C:Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with the enzyme, which are known to have an inhibitory effect on the enzyme. Evidence for the existence of two isocolipases in horse pancreas.
C:Genetics:
A:Gene: GDB:CLPS
A:Cross-references: GDB:127277; OMIM:120105
A:Map position: 6pter-6p21.1
A:Introns: 28/3; 69/3
C:Superfamily: colipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-22/Domain: amino-terminal propeptide #status predicted <APP>
F:23-108/Product: colipase #status experimental <MAT>
F:109-112/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F:34-104,40-56,44-80,45-78,66-86/Disulfide bonds: #status predicted
F:69,72,75,76/Binding site: micellar substrate (Lys, Tyr, Tyr) #status predicted

Query Match 20.3%; Score 135.5; DB 1; Length 112;
Best Local Similarity 28.8%; Pred. No. 6.8e-06;
Matches 32; Conservative 17; Mismatches 51; Indels 11; Gaps 3;

QY 12 LFFELFLRGLSPKYNL--LEKESCIRNQCETCCORAPDNCESHCARKEGSG 68
Db 4 ILILLVSLVAYAAPGRGIINLENGELOMSAQCKNCCOHSSALGLARTSMASEN 63
QY 69 SLCTQVFFGQYRACPLRLNLTCIYSKNEKWL-----SIAYGRCKQIGRK 114
Db 64 SECSKVTLYGIYKPCERGLTC---EGDKTIVGSIITNTNFGICHADGRSK 111

RESULT 3
XLHOA
colipase A precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C:Accession: S53510
R:Rugani, N.; Carrier, F.; Thim, L.; Borgstrom, B.; Sarda, L.
Biochim. Biophys. Acta 1247, 185-194, 1995
A:Title: Lipid binding and activating properties of porcine pancreatic colipase split at the amino-terminal propeptide.
A:Reference number: S53510; MUID:95210311
A:Accession: S53510
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-95 <RUG>
C:Superfamily: colipase

Query Match 20.1%; Score 134; DB 1; Length 95;
Best Local Similarity 34.1%; Pred. No. 8.3e-06;
Matches 29; Conservative 11; Mismatches 37; Indels 8; Gaps 2;

QY 33 LEKESCIRNQCETCCORAPDNCESHCARKEGSGSLCQTQVFFGQYRACPLRLNLTCI 92

Db 11 LDEGELCLNSAQCKNCCQHDHTILSRCAKARENSECSAFTLYGVYKPCERGLTC- 69
QY 93 YSKNEKWL-----SIAYGRCKQIGR 112
Db 70 --EGDKSLVGSITNTNFGICHNVGR 92

RESULT 4
XLHOA
colipase B precursor - horse
N:Alternate names: procolipase B
C:Species: Equus caballus (domestic horse)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 26-Apr-1996
C:Accession: A03165; B90220
R:Bonicel, J.; Couchoud, P.; Foglizzo, E.; Desnuelle, P.; Chapus, C.
Biochim. Biophys. Acta 669, 39-45, 1981
A:Title: Amino acid sequence of horse colipase B.
A:Reference number: A90637; MUID:82046794
A:Accession: A03165
A:Molecule type: protein
A:Residues: 1-96 <BON>
R:Julien, R.; Bechis, G.; Gregoire, J.; Rathelot, J.; Rochat, H.; Sarda, L.
Biochem. Biophys. Res. Commun. 95, 1245-1252, 1980
A:Title: Evidence for the existence of two isocolipases in horse pancreas.
A:Reference number: A90220; MUID:81021166
A:Accession: B90220
A:Molecule type: protein
A:Residues: 1-21, 'E', '23-28, 'T', '30-55 <JUL>
C:Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 stoichiometric complex with the enzyme, which are known to have an inhibitory effect on the enzyme. Evidence for the existence of two isocolipases in horse pancreas.
C:Superfamily: colipase
C:Keywords: lipid digestion; lipid hydrolysis; pancreas
F:1-3/Domain: propeptide #status experimental <PRO>
F:16-96/Product: colipase B #status experimental <MAT>
F:17-87,23-39,27-63,28-61,49-69/Disulfide bonds: #status predicted
F:52,55,58,59/Binding site: micellar substrate (Trp, Tyr, Tyr) #status predicted

Query Match 20.1%; Score 134; DB 1; Length 95;
Best Local Similarity 39.7%; Pred. No. 8.3e-06;
Matches 25; Conservative 6; Mismatches 33; Indels 0; Gaps 0;

QY 33 LEKESCIRNQCETCCORAPDNCESHCARKEGSGSLCQTQVFFGQYRACPLRLNLTCI 92
Db 11 LEAGEICMNSAQCKSECCCHRESSLSLACAKASENSECSAWTLYGVYKPCERGLTCQ 70
QY 93 YSK 95
Db 71 VDK 73

RESULT 5
XLHOA
colipase A precursor - horse
N:Alternate names: procolipase A
C:Species: Equus caballus (domestic horse)
C:Date: 14-Nov-1983 #sequence_revision 04-Dec-1986 #text_change 26-Apr-1996
C:Accession: A03164; A91119; A90220
R:Sternby, B.; Engstrom, A.; Hellman, U.; Vihert, A.M.; Sternby, N.H.; Borgstrom, B.
Biochim. Biophys. Acta 784, 75-80, 1984
A:Title: The primary sequence of human pancreatic colipase.
A:Reference number: A90652; MUID:84104937
A:Accession: A03164
A:Molecule type: protein
A:Residues: 1-96 <STE>
A:Note: Residues 56-59 were positioned by homology; no overlap was obtained for 65-66
R:Pierrot, M.; Astier, J.P.; Astier, M.; Charles, M.; Drenth, J.
Eur. J. Biochem. 123, 347-354, 1982
A:Title: Pancreatic colipase: crystallographic and biochemical aspects.
A:Reference number: A91119; MUID:82186702
A:Accession: A91119
A:Molecule type: protein

A;Accession: A03162
A;Molecule type: protein
A;Residues: 1-95 <STE>
R;Charles, M.; Erlanson, C.; Bianchetta, J.; Joffre, J.; Guidoni, A.; Roversy, M.
Biochim. Biophys. Acta 359, 186-197, 1974
A;Title: The primary structure of porcine colipase II. I. The amino acid sequence.
A;Reference number: A90593; MUID:74290109
A;Accession: A90593
A;Molecule type: protein
A;Residues: 6-91 <CHA>
R;Erlanson, C.; Charles, M.; Astier, M.; Desnuelle, P.
Biochim. Biophys. Acta 359, 198-203, 1974
A;Title: The primary structure of porcine colipase II. II. The disulfide bridges.
A;Reference number: A90594; MUID:74290110
A;Contents: annotation; disulfide bonds
C;Comment: Colipase, a cofactor of triacylglycerol lipase (EC 3.1.1.3), forms a 1:1 s
se the enzyme is washed off by bile salts, which are known to have an inhibitory effe
C;Comment: Residues 6-9 and Arg-92 are considered essential for the function of colip
C;Superfamily: colipase
C;Keywords: lipid digestion; lipid hydrolysis; pancreas
F;1-5/Domain: propeptide #status experimental <PRO>
F;6-95/Product: colipase II #status experimental <MAT>
F;17-87,23-39,27-63,28-61,49-69/Disulfide bonds: #status experimental
F;52,55,58,59/Binding site: micellar substrate (Phe, Tyr, Tyr, Tyr) #status predicted

Query Match 19.2%; Score 128; DB 1; Length 95;
Best Local Similarity 32.9%; Pred. No. 3e-05;
Matches 28; Conservative 11; Mismatches 38; Indels 8; Gaps 2;

QY 33 LELKESIRNQDCFTGCCORAPDNCHSAPEKGSGSLCQTQVFPGYACPLRNITCI 92
| : | : | : |||| | | | | | : | | | |
Db 11 LDEGELCLNSAQKSNCCQHDTILSLLRCALKARENSECSAFTLYGVYYKPCBERGLTC- 69

QY 93 YSKNEKWL-----SIAYGRCKQIGR 112
| : | | : : | : | |
Db 70 --EGDKSLVGSITNTNFGICHNVGR 92

RESULT 8
I51909
colipase precursor - rat
N;Alternate names: procoplipase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Apr-2000
C;Accession: I51909; A34623
R;Payne, R.M.; Sims, H.F.; Jennens, M.L.; Lowe, M.E.
Am. J. Physiol. 266, G914-G921, 1994
A;Title: Rat pancreatic lipase and two related proteins: enzymatic properties and mRNA
A;Reference number: I51909; MUID:94262798
A;Accession: I51909
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-112 <PAY>
A;Cross-references: GB:M58370; NID:g203504; PIDN:AAA20505.1; PID:g203505
R;Wicker, C.; Puigserver, A.
Biochem. Biophys. Res. Commun. 167, 130-136, 1990
A;Title: Rat pancreatic colipase mRNA: nucleotide sequence of a cDNA clone and nutrit
A;Reference number: A34623; MUID:90179738
A;Accession: A34623
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-17,'V',19-112 <WIC>
A;Cross-references: GB:M33333; NID:g203502; PIDN:AAA40943.1; PID:g203503
C;Superfamily: colipase
C;Keywords: lipid digestion; lipid hydrolysis; pancreas
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-112/Product: colipase #status predicted <MAT>

Query Match 18.3%; Score 122; DB 2; Length 112;
Best Local Similarity 31.9%; Pred. No. 0.00012;
Matches 29; Conservative 8; Mismatches 38; Indels 16; Gaps 2;

```
QY 33 LELKESIRNQDCTGCCORAPDNCEHCAEKGSGSLCQTVFFGQYRACPLRLNLTIC- 91
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 28 LEDGEICVNSMOCKSRCCQHDHILGATRTHKAMENSECSPKTLGYIYRCPCERGLTCE 87
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
QY 92 -----IYSKNEKWLSTAYGRCKQIGROK 114
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 88 GDRSIIGAITNTN-----YGVCLDSTRSK 111
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 9
S08220
platelet-derived growth factor chain A precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
C:Accession: S08220
R:Bejcek, B.E.; Li, D.Y.; Deuel, T.F.
Nucleic Acids Res. 18, 680, 1990
A:Title: Nucleotide sequence of a cDNA clone of Xenopus platelet-derived growth factor A
A:Reference number: S08220; MUID: 90175018
A:Accession: S08220
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-215 <BEJ>
A:Cross-references: EMBL:X17545; NID:g64973; PIDN:CAA35583.1; PID:g64974
C:Superfamily: platelet-derived growth factor
C:Keywords: alternative splicing; growth factor
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-91/Domain: propeptide #status predicted <PRO>
F:92-215/Product: platelet-derived growth factor chain A #status predicted <MAT>

Query Match 12.5%; Score 83.5; DB 2; Length 215;
Best Local Similarity 22.8%; Pred. No. 0.74; Mismatches 48; Indels 19; Gaps 5;
Matches 26; Conservative 21;

QY 17 LFLLTRGLSPTKYNLELKESIRNQDCTGCCORAPDNCE-----SHCAEK-----G 65
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 107 IYEIPRSDIPTSANFL-IWPCVEVKRC-TGCCNTSSVKQCPSRTHRSVKVAKVEYVR 164
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

QY 66 SEGSLCQTVFFGQYRACPLRLNLTICYSKNEKWLSTAYGRCKQIGROKLAKKM 119
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 165 KKPRLKVLVRLHELECTCTANSNDYREET-----GRTRETKRKRKKL 212
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 10
T21762
hypothetical protein F35C5.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21762; T22937
R:Sims, M.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19469
A:Accession: T21762
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-126 <WTL>
A:Cross-references: EMBL:Z81076; NID:e1061694; PIDN:CAB03062.1; GSPDB:GN000020; CESP:F35C5
A:Experimental source: clone F35C5
R:Smey, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19639
A:Accession: T22937
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-126 <WIL2>
A:Cross-references: EMBL:Z81556; PIDN:CAB04528.1; GSPDB:GN000020; CESP:F35C5.12
A:Experimental source: clone F58G1
C:Genetics:
A:Gene: CESP:F35C5.12
A:Map position: 2
A:Introns: 18/2; 52/2; 81/3
```

```
Query Match 12.1%; Score 80.5; DB 2; Length 136;
Best Local Similarity 23.0%; Pred. No. 0.95;
Matches 31; Conservative 13; Mismatches 44; Indels 47; Gaps 6;

QY 9 LFLFLFFFLTRGLSPTKYNLELKESIRNQDCTGCC-----IRNODCTGC----- 49
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 1 MLFL-FFAFL-----SLQP-----FVRGACTFLDHOACDEVCKTDSFWYGHCHNGWDGFN 50
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

QY 50 -----CORAPDNCEHCAEKGSGSLC-QTVFFGQYRACPLRLNLTIC 91
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 51 FSCKYEYITPLDGSICETROMACSEKDKGSGEGGFCFPQLDSHKSLRTACECFKKLQV 110
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

QY 92 IYSKNEKWLSTAYGR 106
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 111 LRRKRSIQRSYKR 125
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 11
T26972
hypothetical protein Y47H9C.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T26972
R:Harris, B.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20293
A:Accession: T26972
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1111 <WIL>
A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN000019; CESP:Y47H9C.4
A:Experimental source: clone Y47H9C
C:Genetics:
A:Gene: CESP:Y47H9C.4
A:Map position: 1
A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match 12.0%; Score 80; DB 2; Length 1111;
Best Local Similarity 31.2%; Pred. No. 5.3;
Matches 24; Conservative 8; Mismatches 27; Indels 18; Gaps 6;

QY 44 DCETGC-CORAP--DNCESHC-AEKGSGSLCQTVQV---FFGQ--YRACPLRLNLTICYS 94
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 244 ECKFECCNGATCDNTNGCKICKSGYHGALCENECVSGFFGSGCTQKCDCLNNQNCDS 303
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

QY 95 KNEKWLSTAYGRCKQIG 111
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 304 S-----GECKCIG 311
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 12
I51550
platelet-derived growth factor A chain long form precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
C:Accession: I51550
R:Mercola, M.; Melton, D.A.; Stiles, C.D.
Science 241, 1223-1225, 1988
A:Title: Platelet-derived growth factor A chain is maternally encoded in Xenopus embr
A:Reference number: I51550; MUID: 88321676
A:Accession: I51550
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-226 <MER>
A:Cross-references: GB:M23237; NID:g214648; PIDN:AAA49927.1; PID:g214649
C:Superfamily: platelet-derived growth factor

Query Match 11.9%; Score 79; DB 2; Length 226;
```

Best Local Similarity 21.8%; Pred. No. 2;
Matches 26; Conservative 22; Mismatches 53; Indels 18; Gaps 5;

QY 17 LFLTRGSLPTKYNLLEKESCRNODCEGCGCORAPDNC-----SHCAEK-----G 65
Db 107 IYETPRQIDPTSANFL-IWPCVVEVRK-TGCCNTSVKCPRIHRSVKVAKVEYVR 164
QY 66 SEGSLCQTQVFFGQYRACPCPLRNLTCTIYSKNEKWL-----SIAYGRCQKIGRQKLAKKM 119
Db 165 KKPALKVLRLERLEHCTCTANSNSDYREETGFTSPALVLTGRTRETGKKOKRKL 223

RESULT 13
S18188
notch protein homolog - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999
C:Accession: S18188
R;Weinmaster, G.; Roberts, V.J.; Lemke, G.
Development 113, 199-205, 1991
A:Title: A homolog of Drosophila Notch expressed during mammalian development.
A:Reference number: S18188; MUID:92111383
A:Accession: S18188
A:Molecule type: mRNA
A:Residues: 1-2531 <WEI>
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:967-1018/Domain: EGF homology <EGF1>
F:1025-1056/Domain: EGF homology <EGF2>
F:1233-1264/Domain: EGF homology <EGF3>
F:1917-1949/Domain: ankyrin repeat homology <AN1>
F:1950-1982/Domain: ankyrin repeat homology <AN2>
F:1984-2016/Domain: ankyrin repeat homology <AN3>
F:2017-2049/Domain: ankyrin repeat homology <AN4>
F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 11.8%; Score 78.5; DB 2; Length 2531;
Best Local Similarity 29.5%; Pred. No. 13;
Matches 23; Conservative 11; Mismatches 35; Indels 9; Gaps 5;

QY 15 FFLFLTRGSLPT-KYNLLEKESCRNODCEGCGCORAPDNCESHCAEKSGSLCQT 73
Db 624 YYLCLLKGITGPNCETNL-----DDCASN-PCDSGTCLDKIDGYECAC-EPGYTGSMCNV 677

QY 74 QVFFGQYRACPCPLRNLT 91
Db 678 NI--DECAGSPCHNGGTC 693

RESULT 14
S43910
gibberellin-regulated protein RSI-1 precursor - tomato
N;Alternate names: GAST1 protein homolog
C:Species: Lycopersicon esculentum (tomato)
C:Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 24-Sep-1999
C:Accession: S43910; S43909
R;Taylor, B.H.; Scheuring, C.F.
Mol. Gen. Genet. 243, 148-157, 1994
A:Title: A molecular marker for lateral root initiation: the RSI-1 gene of tomato (Lycopersicon)
A:Reference number: S43909; MUID:94232181
A:Accession: S43910
A:Molecule type: DNA
A:Residues: 1-96 <TAY>
A:Cross-references: EMBL:L22189; NID:g405586; PIDN:AAA20130.1; PID:g405587
A:Accession: S43909
A:Molecule type: mRNA
A:Residues: 1-96 <TAY>
A:Cross-references: EMBL:L22188; NID:g405584; PIDN:AAA20129.1; PID:g405585
C:Genetics:
A:Gene: RSI-1
A:Introns: 28/3; 37/1
C:Superfamily: gibberellin-regulated protein GASA2

F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-96/Product: gibberellin-regulated protein RSI-1 #status predicted <MAT>

Query Match 11.7%; Score 78; DB 2; Length 96;
Best Local Similarity 28.4%; Pred. No. 1.3;
Matches 25; Conservative 9; Mismatches 32; Indels 22; Gaps 4;

QY 10 LFLPFFFLTRGSLPTKYNLLEKESCRNODCEGCGCORAPDN-----CESH 60
Db 11 LLLISMFLLILTFNSV-VEGYNKL-----RPTDCKPCTYRCSATSHKKRPFQCKC 62

QY 61 CAEKSGESGLCQTQVFFGQYRACPCPLRN 88
Db 63 CAT-----CLCVPGVYGKQSCPCYNN 85

RESULT 15
A57480
tubulointerstitial nephritis antigen precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 05-Nov-1999
C:Accession: A57480
R;Nelson, T.R.; Charonis, A.S.; McIvor, R.S.; Butkowski, R.J.
J. Biol. Chem. 270, 16265-16270, 1995
A:Title: Identification of a cDNA encoding tubulointerstitial nephritis antigen.
A:Reference number: A57480; MUID:95332335
A:Accession: A57480
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-474 <NEL>
A:Cross-references: GB:U24270; NID:g818410; PIDN:AAC48477.1; PID:g818411
C:Superfamily: somatomedin B homology
C:Keywords: basement membrane; extracellular matrix; glycoprotein
F;51-104/Domain: somatomedin B homology <SBH2>

Query Match 11.6%; Score 77.5; DB 2; Length 474;
Best Local Similarity 25.0%; Pred. No. 4.8;
Matches 28; Conservative 18; Mismatches 43; Indels 23; Gaps 7;

QY 6 WLLLLFLFFFLF-----LLTRGSLPTKYNLLE---LKESCIRNQDCET-GCC 50
Db 2 WTGYKFFIFFYLAKEIWTPEKQKQVDLASYSSRNHSILEGTRFKRSIFOGGYCRSLGCC 61

QY 51 QRAPDNCESHAEGSGSLCQTQVFFGQYRA--CPCLRNLTCTIYSKNEKWL 100
Db 62 EGRNDNCVTQFYE---ANALCYCDKFCERENSDCCPDYKSF-C--QEEKEWL 107

Search completed: March 24, 2001, 13:23:16
Job time: 106 sec

QY 79 QYRACPLRLNLTCTYKNEKW--LSIAYGRCKQIGKQKLA 117
 Db 598 ECKMPECHTIPCLLSPWSEWSDCSVTGCKGMRT-RQRLMK 637

RESULT 2

US-08-313-288B-12
 ; Sequence 12, Application US/08313288B
 ; Patent No. 5750502
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M. and AviHu Klar
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313.288B
 ; FILING DATE: January 5, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0526
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 802 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US93-0313-288B-12

Query Match 11.4%; Score 76; DB 1; Length 802;
 Best Local Similarity 21.8%; Pred. No. 5.7;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;

QY 34 ELKESCI RNQDCETGCCQRAP-----DNCESHCAE-----KGSEGLCQTQVFFG 78
 Db 538 ETEKCI VNEECSPSCLVTWGEWDECSASCTGMRKRHRMIKMTPADGSMCKAETTEA 597

RESULT 3

PCT-US93-03164-12
 ; Sequence 12, Application PC/TUS9303164
 ; Patent No. 5750502
 ; GENERAL INFORMATION:
 ; APPLICANT: Jessell, Thomas M
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York

STATE: New York
 COUNTRY: USA
 ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/03164
 ; FILING DATE: 19930402
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 40028
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 977-9550
 ; TELEFAX: (212) 664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 802 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US93-03164-12

Query Match 11.4%; Score 76; DB 4; Length 802;
 Best Local Similarity 21.8%; Pred. No. 5.7;
 Matches 22; Conservative 17; Mismatches 44; Indels 18; Gaps 4;

QY 34 ELKESCI RNQDCETGCCQRAP-----DNCESHCAE-----KGSEGLCQTQVFFG 78
 Db 538 ETEKCI VNEECSPSCLVTWGEWDECSASCTGMRKRHRMIKMTPADGSMCKAETTEA 597

RESULT 4

US-08-597-545-2
 ; Sequence 2, Application US/08597545
 ; Patent No. 5580738
 ; GENERAL INFORMATION:
 ; APPLICANT: LABORDA, Jorge
 ; TITLE OF INVENTION: Delta-Like Gene Expressed In
 ; TITLE OF INVENTION: Neuroendocrine Tumors
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/597.545
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/989,537
 ; FILING DATE: 11-DEC-1992
 ; ATTORNEY/AGENT INFORMATION:

; INFORMATION FOR SEQ ID NO: 1:

03-08-437-133-1

53 LCPKCVTAPGCVNGVC-KEPWOC---ICKDGDGKCFCEIDV-----RACTSTPCANNGTC 103


```

; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-03164-10

Query Match 11.0%; Score 73; DB 4; Length 807;
Best Local Similarity 20.4%; Pred. No. 11;
Matches 21; Conservative 19; Mismatches 45; Indels 18; Gaps 4;

Qy 34 ELKESIRNODTCCQRAP-----DNCESHCA-----EKGSESLCQTQVFFG 78
Db 543 EETKCTVNEECSPSLTENGWDGSDATCGMGKRRHVMKMSPADGSMCKRAETSOA 602

Qy 79 QYRACPLRLNLCIYSKNEKW--LSIAYGRCKIGRKLAKKM 119
Db 603 EKCMMECHTIFCLLSPWSEWSDCVTCGKGMRT-RQRLKSL 644

RESULT 13
US-08-036-555B-40
; Sequence 40, Application US/08036555B
; Patent No. 5530109
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,555B
; FILING DATE: 24-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.

; NAME: White, John P
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 638-3884
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-036-555B-40

Query Match 10.9%; Score 72.5; DB 1; Length 39;
Best Local Similarity 31.9%; Pred. No. 0.57;
Matches 15; Conservative 2; Mismatches 19; Indels 11; Gaps 2;

Qy 45 CETGCCORAPDNCESHAKEGSESLCQTQVFFGOYRACPLRLNLC 91
Db 1 CTCGCCCKCRRTTACRCA--GAAGGTCTT-----CTCCTTCTC 36

RESULT 14
US-08-469-569-40
; Sequence 40, Application US/08469569
; Patent No. 5606032
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew; Stroobant, Paul;
; APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Maio Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,569
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
```


us-09-215-435-179.ra1

Sat Mar 24 13:26:25 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:24:20 ; Search time 25.25 Seconds
(without alignments)
126.618 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MSAKLGFLRRFFIFCSLNT.....NFKLKIEREVACVAKYKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	154	28.6	110	IBP_CARCR	P00993 caretta car
2	134	24.9	352	AMB_BOVIN	P00978 bos taurus
3	130	24.1	62	IP52_ANESU	P10280 anemonia su
4	130	24.1	83	ELAC_MACEU	O62845 macropus eu
5	129.5	24.0	352	AMB_HUMAN	P02760 homo sapien
6	127	23.6	65	IVB3_VIPAA	P00992 vipera ammo
7	127	23.6	123	IATR_HORSE	P04365 equus cabal
8	127	23.6	123	IATR_SHEEP	P13371 ovis aries
9	126	23.4	67	IBFC_BOVIN	P00976 bos taurus
10	126	23.4	102	ELAC_TRIVU	Q29143 trichosurus
11	124	23.0	235	TFP2_HUMAN	P48307 homo sapien
12	124	23.0	349	AMB_MOUSE	O07456 mus musculu
13	123	22.8	60	IVBL_DENPO	P00979 dendroaspis
14	122	22.6	349	AMB_RAT	Q64240 rattus norv
15	121	22.4	300	TFPI_RABIT	P19761 oryctolagus
16	120	22.3	304	TFPI_MACMU	Q28864 macaca mula
17	118	21.9	62	IVBT_ERIMA	P24541 eristocophi
18	118	21.9	337	AMB_PIG	P04366 sus scrofa
19	118	21.9	346	AMB_MERUN	Q62377 meriones un
20	117	21.7	57	SBPI_SARBU	P26328 sarcophaga
21	117	21.7	302	TFPI_RAT	Q02445 rattus norv
22	116	21.5	304	TFPI_HUMAN	P10646 homo sapien
23	115	21.3	59	IVBL_DENAN	P00980 dendroaspis
24	113	21.0	349	AMB_MESAU	O60559 mesocricetu
25	112.5	20.9	355	AMB_PLEPL	P36992 pleuronecte
26	111	20.6	57	IVBC_NAJNA	P19859 naja naja
27	110	20.4	2944	TXCA_HUMAN	Q02388 homo sapien
28	109	20.2	60	TXCA_DENAN	P81658 dendroaspis
29	109	20.2	164	TKD1_BOVIN	Q28201 bos taurus
30	108.5	20.1	765	APP2_RAT	P15943 rattus norv
31	108	20.0	62	ISC2_BOMMO	P10832 bombyx mori
32	108	20.0	65	IVBI_BUNFA	P25660 bungarus fa
33	107.5	19.9	763	APP2_HUMAN	Q06481 homo sapien

```
34 106 19.7 62 1 ISC1_BOMMO
35 106 19.7 64 1 KIB9_HUMAN
36 105.5 19.6 1416 1 YN81_CAEEL
37 105 19.5 63 1 ICS3_BOMMO
38 105 19.5 3176 1 CA36_HUMAN
39 103 19.1 57 1 IVBK_DENAN
40 103 19.1 57 1 IVBT_NAJNA
41 103 19.1 60 1 IVB2_DABRU
42 103 19.1 265 1 TKD1_SHEEP
43 102 18.9 60 1 IBPS_BOVIN
44 102 18.9 61 1 IVBL_VIPAA
45 102 18.9 122 1 UPTI_PIG
```

```
P10831 bombyx mori
P49223 homo sapien
O03610 caenorhabdi
P07481 bombyx mori
P12111 homo sapien
P00982 dendroaspis
P20229 naja naja
P00990 daboia russ
Q29428 ovis aries
P00975 bos taurus
P00991 vipera ammo
Q29100 sus scrofa
```

ALIGNMENTS

```
RESULT 1
IBP_CARCR
ID IBP_CARCR STANDARD; PRT; 110 AA.
AC P00993;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CHELONIANIN (BASIC PROTEASE INHIBITOR) (RTPI).
OC Caretta caretta (Loggerhead).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Chelonioidae; Cheloniidae; Caretta.
RN [1]
RP TISSUE=EGG WHITE;
RC Kato I., Tominaga N.;
RT "Trypsin-subtilisin inhibitor from red sea turtle eggwhite consists
of two tandem domains -- one Kunitz -- one of a new family.";
RL Fed. Proc. 38:832-832(1979).
CC -I- MISCELLANEOUS: THIS INHIBITOR, ISOLATED FROM EGG WHITE, CONSISTS
OF TWO NONHOMOLOGOUS DOMAINS.
CC -I- SIMILARITY: THE FIRST DOMAIN BELONGS TO THE BPTI/KUNITZ FAMILY OF
INHIBITORS.
CC -I- SIMILARITY: THE SECOND DOMAIN IS A WAP-TYPE 'FOUR-DISULFIDE CORE'
DOMAIN.
CC -I- CAUTION: AS THE PAPER ONLY INDICATES THE SPECIES AS "RED SEA
TURTLE", THE SPECIES INDICATED HERE IS THEREFORE AN INFERENCE.
CC PIR; A01224; TITROR.
DR HSSP; P00974; 1BRB.
DR INTERPRO; IPR002221; -.
DR INTERPRO; IPR002223; -.
DR PFAM; PF00014; Kunitz_Bpti; 1.
DR PFAM; PF00095; wap; 1.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00317; 4-DISULFIDE_CORE; 1.
DR PROSITE; PS0289; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0289; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor.
FT DOMAIN 1 58 INHIBIT TRYPSIN.
FT DOMAIN 59 110 INHIBIT SUBTILISIN.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT ACT_SITE 18 19 REACTIVE BOND (TRYPSIN).
SQ SEQUENCE 110 AA; 11916 MW; 269436243813418E CRC64;
```

Query Match 28.6%; Score 154; DB 1; Length 110;

Best Local Similarity 46.9%; Pred. No. 9.4e-11; Mismatches 23; Indels 4; Gaps 1;

Matches 30; Conservative 7; YQ 35 GDLKDPCKLMNFGSCYEYHFRFYNYRTSKRCFTVFGCGNGLNNFKLKIEREVACVAK 94

Db 2 GDRKIDICLPEQPCPKGRIPRYFNPNASRCESFIYGGCKGNKNNFKTAE-----CVRA 57

YQ 95 YKPP 98

Db 58 CRPP 61

[illegible]

RT RC TISSUE=LIVER;
RX MEDLINE=90336621; PubMed=1696200;
RA Diarra-Mehrpour M., Bourguignon J., Sessboue R., Salier J.P.,
RA Leveillard T., Martin J.P.; the human inter-alpha-trypsin inhibitor
RT "Structural analysis of the human inter-alpha-trypsin inhibitor
RT light-chain gene";
RL Eur. J. Biochem. 191:131-139(1990).
RN [4]
RN SEQUENCE OF 1-220 FROM N.A.
RP MEDLINE=86312901; PubMed=2428011;
RX Traboni C., Cortese R.;
RA "Sequence of a full length cDNA coding for human protein HC (alpha 1
RT microglobulin).";
RL Nucleic Acids Res. 14:6340-6340(1986).
RN [5]
RN SEQUENCE OF 20-202 (INDIVIDUAL WITH TUBULAR PROTEINURIA).
RX MEDLINE=84126849; PubMed=6198962;
RA Lopez C., Grubb A.O., Mendez E.;
RT "The complete amino acid sequence of human complex-forming
RT glycoprotein heterogeneous in charge (protein HC) from one
RT individual";
RL Arch. Biochem. Biophys. 228:544-554(1984).
RN [6]
RN SEQUENCE OF 20-198 (VARIANT).
RA Lopez C., Grubb A.O., Mendez E.;
RT "Human protein HC displays variability in its carboxyl-terminal amino
RT acid sequence";
RL FEBS Lett. 144:349-353(1982).
RN [7]
RN SEQUENCE OF 20-198 (PATIENTS WITH TUBULAR PROTEINURIA).
RX MEDLINE=81184038; PubMed=6164372;
RA Takagi T., Takagi K., Kawai T.;
RT "Complete amino acid sequence of human alpha 1-microglobulin";
RL Biochem. Biophys. Res. Commun. 98:997-1001(1981).
RN [8]
RN SEQUENCE OF 206-350.
RX MEDLINE=85225968; PubMed=2408638;
RA Reisinger P., Hochstrasser K., Albrecht G.J., Lempart K., Salier J.P.;
RT "Human inter-alpha-trypsin inhibitor: localization of the Kunitz-type
RT domains in the N-terminal part of the molecule and their release by a
RT trypsin-like proteinase";
RL Biol. Chem. Hoppe-Seyler 366:479-483(1985).
RN [9]
RN CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=82074265; PubMed=6171497;
RA Hochstrasser K., Schoenberger O.L., Rossmann I., Wächter E.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of
RT the inter-alpha-trypsin inhibitor, V. Attachments of carbohydrates in
RT the human urinary trypsin inhibitor isolated by affinity
RT chromatography";
RL Hoppe-Seyler's Z. Physiol. Chem. 362:1357-1362(1981).
RN [10]
RN INHIBITORY SITE.
RX MEDLINE=85225940; PubMed=3890890;
RA Morli M., Travis J.;
RT "The reactive site of human inter-alpha-trypsin inhibitor is in the
RT amino-terminal half of the protein";
RL Biol. Chem. Hoppe-Seyler 366:19-21(1985).
RN [11]
RN STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=90306345; PubMed=1694784;
RA Escibano J., Lopez-Otin C., Hierpe A., Grubb A., Mendez E.;
RT "Location and characterization of the three carbohydrate prosthetic
RT groups of human protein HC";
RL FEBS Lett. 266:167-170(1990).
RN [12]
RN CHROMOPHORE.
RC TISSUE=URINE;
RX MEDLINE=91340714; PubMed=1714898;
RA Escibano J., Grubb A., Calero M., Mendez E.;
RT "The protein HC chromophore is linked to the cysteine residue at
RT position 34 of the polypeptide chain by a reduction-resistant bond

RT and causes the charge heterogeneity of protein HC";
RL J. Biol. Chem. 266:15758-15763(1991).
RN [13]
RN SEQUENCE OF 206-219, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
RC TISSUE=PLASMA;
RX MEDLINE=9422087; PubMed=7513643;
RA Morelle W., Capon C., Balduyck M., Sautiere P., Kouach M.,
RA Michalski C., Fournet B., Mizon J.;
RT "Chondroitin sulphate covalently cross-links the three polypeptide
RT chains of inter-alpha-trypsin inhibitor";
RL Eur. J. Biochem. 221:881-888(1994).
RN [14]
RN SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC2.
RX MEDLINE=93232026; PubMed=7682553;
RA Enghild J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
RA Pizzo S.V., Hefta S.A.;
RT "Presence of the protein-glycosaminoglycan-protein covalent cross-link
RT in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
RT 2/bikunin";
RL J. Biol. Chem. 268:8711-8716(1993).
RN [15]
RN SEQUENCE OF 206-223, AND CROSS-LINK SITE TO HC3.
RX MEDLINE=91093267; PubMed=1898736;
RA Enghild J.J., Salvesen G., Hefta S.A., Thøgersen I.B.,
RA Rutherford S., Pizzo S.V.;
RT "Chondroitin 4-sulfate covalently cross-links the chains of the human
RT blood protein pre-alpha-inhibitor";
RL J. Biol. Chem. 266:747-751(1991).
RN [16]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 230-339.
RX MEDLINE=9827321; PubMed=9566199;
RA Xu Y., Carr P.D., Guss J.M., Ollis D.L.;
RT "The crystal structure of bikunin from the inter-alpha-inhibitor
RT complex: a serine protease inhibitor with two Kunitz domains";
RL J. Mol. Biol. 276:953-966(1998).
CC -1- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN.
CC -1- FUNCTION: INTER-ALPHA-TRYPsin INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPsin, PLASMIN, AND LYSOSOMAL GRANULOCYTIC
CC ELASTASE. ADDITIONAL PROTEOLYTIC PROCESSING IN THE KIDNEY AND/OR
CC URINE CAN PRODUCE FURTHER AMINO- AND CARBOXYL-END MODIFICATIONS
CC IN ITS SEQUENCE.
CC -1- SUBUNIT: INTER-ALPHA-TRYPsin INHIBITOR CONSIST OF A LIGHT CHAIN
CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
CC -1- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC -1- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE.
CC CROSS-LINKING BETWEEN THE DIFFERENT COMPONENTS.
CC -1- MISCELLANEOUS: IN VITRO, THE FIRST TWELVE RESIDUES OF THE AMINO
CC END OF THE INHIBITOR APPEAR TO HAVE A REACTIVE SITE CAPABLE OF
CC INHIBITING THE ACTIVITY OF A NUMBER OF ENZYMES. ITS IN VIVO
CC FUNCTION IS NOT KNOWN.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE BPTI/KUNITZ
CC FAMILY OF INHIBITORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54816; CAA38585.1;
CC EMBL; X54817; CAA38585.1; JOINED.
CC EMBL; X54818; CAA38585.1; JOINED.
CC EMBL; X04225; CAA27803.1;
CC EMBL; M88249; AAA59196.1;


```

DR EMBL: M88165; AAA59196.1; JOINED.
DR EMBL: M88243; AAA59196.1; JOINED.
DR EMBL: M88244; AAA59196.1; JOINED.
DR EMBL: M88246; AAA59196.1; JOINED.
DR EMBL: M88247; AAA59196.1; JOINED.
DR EMBL: X04494; CAA28182.1; -.
DR EMBL: X34817; CAA38586.1; -.
DR PIR: A03217; HCHU.
DR PIR: A25303; A25303.
DR PIR: S13433; S13433.
DR PIR: S10717; S10717.
DR PDB: 1BK; 16-MAR-99.
DR SWISS-2DPAGE; P02760; HUMAN.
DR MIM: 176870; -.
DR INTERPRO; IPR000566; -.
DR INTERPRO; IPR002223; -.
DR INTERPRO; IPR002345; -.
DR PFAM; PF00014; Kunitz_BPTI; 2.
DR PRINTS; PR00179; LIPOCALIN; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
KW Lipocalin; 3D-structure.

Query Match 24.0%; Score 129.5; DB 1; Length 352;
Best Local Similarity 41.4%; Pred. No. 1.9e-07;
Matches 29; Conservative 9; Mismatches 25; Indels 7; Gaps 2;

QY 24 GG--VNKIAEKIGDLKPKCKLDMNFGSCVEYHFRFYNTSKRCETFFVSGCGNGLNPF 81
| | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 217 GGQLVTEVTKK-----EDSQLGYSAGPCMGMTSRFYNGTSMACETFFQGGCGMGNNF 271
| | : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 82 KLKIEREVAC 91
| | : | |
DB 272 VTEKECLQTC 281

RESULT 6
IVB3_VIPAA
ID IVB3_VIPAA STANDARD; PRT; 65 AA.
AC P00992;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE VENOM BASIC PROTEASE INHIBITOR III (VENOM CHYMOTRYPSIN INHIBITOR).
OS Vipera ammodytes ammodytes (Western sand viper).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Viperinae; Vipera.
RN [1]
RP SEQUENCE.
RC TISSUE=VENOM;
RA Ritonja A., Meloun B., Gubensek F.;
RT "The primary structure of Vipera ammodytes venom chymotrypsin inhibitor."
RL Biochim. Biophys. Acta 746:138-145(1983).
CC -!- FUNCTION: THIS PROTEIN INHIBITS CHYMOTRYPSIN.
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR; A01223; TIVIVC.
DR HSSP; P31713; LSHP.
DR INTERPRO; IPR002223; -.
DR PFAM; PF00014; Kunitz_BPTI; 1.
DR PRINTS; PR00759; BASICPTASE.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.
KW Venom; Serine protease inhibitor.
FT DISULFID 7 57
FT DISULFID 16 40
FT DISULFID 32 53
FT ACT_SITE 17 18 REACTIVE BOND (BY SIMILARITY).

```

```

SQ SEQUENCE 65 AA; 7556 MW; 9D526F8E3BF7CC57 CRC64;

Query Match 23.6%; Score 127; DB 1; Length 65;
Best Local Similarity 42.4%; Pred. No. 6.3e-08;
Matches 25; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 41 CKLDMNFGSCVEYHFRFYNTSKRCETFFVSGCGNGLNPFKLKIEREVACVAKYKPPR 99
| | : | | : : : : : : : : : : : : : : : : : : : : :
DB 7 CYLPADPGRCCLAYMPFYNPASNKCEKFIYGGCRGNANNEKTDWDECRHTCVASSIQPR 65

RESULT 7
IATR_HORSE
ID IATR_HORSE STANDARD; PRT; 123 AA.
AC P04365;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE INTER-ALPHA-TRYPsin INHIBITOR (ITI) (HI-14) (INHIBITORY FRAGMENT OF ITI) (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
RN [1]
RP SEQUENCE.
RX MEDLINE-85225967; PubMed-2408637;
RA Hochstrasser K., Wächter E., Albrecht G.J., Reisinger P.;
RT "Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-alpha-trypsin inhibitor, X. The amino-acid sequences of the trypsin-released inhibitors from horse and pig inter-alpha-trypsin inhibitors."
RL Biol. Chem. Hoppe-Seyler 366:473-478(1985).
CC -!- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS. WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND NOT AT ALL WITH PANCREATIC ELASTASE.
CC -!- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR; A01210; TIHOBI.
DR HSSP; P10646; IADZ.
DR INTERPRO; IPR002223; -.
DR PFAM; PF00014; Kunitz_BPTI; 2.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT NON_TER 1 1
FT DOMAIN 1 56 I.
FT DOMAIN 57 123 II.
FT DISULFID 5 55
FT DISULFID 14 38
FT DISULFID 30 51
FT DISULFID 61 111
FT DISULFID 70 94
FT DISULFID 86 107
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN, ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPSIN).
FT CARBOHYD 24 24 N-LINKED (GLCNAC...).
FT NON_TER 123 123
SQ SEQUENCE 123 AA; 13510 MW; CE1A912077441ID5 CRC64;

Query Match 23.6%; Score 127; DB 1; Length 123;
Best Local Similarity 48.1%; Pred. No. 1.2e-07;
Matches 26; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 38 KDPCKLDMNFGSCVEYHFRFYNTSKRCETFFVSGCGNGLNPFKLKIEREVAC 91
| | : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 2 EDSCLDHAQPCGLMISRYFYNGTSMACETFOYGGCLGNNGNFFASQKECLQTC 55

```
RESULT 8
IATR_SHEEP STANDARD; PRT; 123 AA.
AC PI3371;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1994 (Rel. 14, Last annotation update)
DE INTER-ALPHA-TRYPsin INHIBITOR (ITI) (GIK-14) (INHIBITORY FRAGMENT OF
DE ITI) (FRAGMENT).
OS Ovis aries (Sheep), and Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
RN [1]
RN SEQUENCE.
RP SPECIES=SHEEP;
RC MEDLINE=87299012; PubMed=2441725;
RA Rasp G., Hochstrasser K., Wachter E., Reisinger P.W.M.;
RT "The amino-acid sequence of the trypsin-released inhibitor from sheep
inter-alpha-trypsin inhibitor".
RL Biol. Chem. Hoppe-Seyler 368:727-731(1987).
RN [2]
RN SEQUENCE.
RP SPECIES=C.HIRCUS;
RC MEDLINE=90105540; PubMed=2481505;
RA Rasp G., Hochstrasser K., Gerl C., Wachter E.;
RT "Primary structure of a proteinase inhibitor released from goat serum
inter-alpha-trypsin inhibitor".
RL Biochim. Biophys. Acta 999:335-337(1989).
CC -!- FUNCTION: THIS INHIBITORY FRAGMENT, RELEASED FROM NATIVE ITI AFTER
LIMITED PROTEOLYSIS WITH TRYPSIN, CONTAINS TWO HOMOLOGOUS DOMAINS.
WHEREAS THE SECOND DOMAIN IS A STRONG INHIBITOR OF TRYPSIN, THE
FIRST DOMAIN INTERACTS WEAKLY WITH PMN-GRANULOCYTIC ELASTASE AND
NOT AT ALL WITH PANCREATIC ELASTASE.
CC -!- MISCELLANEOUS: THE AMINO ACID AT POSITION P2' (17) APPEARS TO
DETERMINE THE SPECIFICITY OF THE INHIBITION OF DOMAIN I.
CC INHIBITORS WITH METHIONINE IN THIS POSITION INTERACT WEAKLY WITH
CHYMOTRYPSIN AND ELASTASE; THOSE WITH LEUCINE INTERACT STRONGLY.
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR: A29652; A29652.
DR HSSP: P10646; IADZ.
DR INTERPRO: IPR002223; -.
DR PFAM: PF00014; Kunitz_BPTI; 2.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 2.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 2.
KW Plasma; Glycoprotein; Serine protease inhibitor; Repeat.
FT NON_TER 1 56 I.
FT DOMAIN 57 123 II.
FT DISULFID 5 55 BY SIMILARITY.
FT DISULFID 14 38 BY SIMILARITY.
FT DISULFID 30 51 BY SIMILARITY.
FT DISULFID 61 111 BY SIMILARITY.
FT DISULFID 70 94 BY SIMILARITY.
FT DISULFID 86 107 BY SIMILARITY.
FT ACT_SITE 15 16 INHIBITORY SITE (P1) (CHYMOTRYPSIN,
ELASTASE).
FT ACT_SITE 71 72 INHIBITORY SITE (P1) (TRYPSIN).
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .).
FT NON_TER 123 123
SQ SEQUENCE 123 AA; 13686 MW; 295038173F2D2D2D1 CRC64;
```

Query Match 23.6%; Score 127; DB 1; Length 123;
Best Local Similarity 46.3%; Pred. No. 1.2e-07;
Matches 25; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 38 KDCKLDNFGSCYEVHFRYFNRTSKRCETFFVSGCNGNLFNFKLIEREVACVAKYKPP 91

Db 2 EDSCLGYSGQPCGLMISRYFYNGTSMACETFOYGGCLGNNGNFFASQKECLQTC 55

```
RESULT 9
IBPC_BOVIN STANDARD; PRT; 67 AA.
AC P00976;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE COLOSTRUM TRYPSIN INHIBITOR (COLOSTRUM BPI).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
RN [1]
RN SEQUENCE.
RP Cechova D., Jonakova V., Sorm F.;
RA "Primary structure of trypsin inhibitor from cow colostrum (component
B2)".
RT Collect. Czech. Chem. Commun. 36:3342-3357(1971).
RN [2]
RN DISULFIDE BONDS.
RA Cechova D., Ber E.;
RT "Disulfide bonds of trypsin inhibitor from cow colostrum."
RL Collect. Czech. Chem. Commun. 39:680-688(1974).
RN [3]
RN CHARACTERIZATION.
RP Cechova D., Muszynska G.;
RA "Role of lysine 18 in active center of cow colostrum trypsin
inhibitor."
RT FEBS Lett. 8:84-86(1970).
CC -!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
DR PIR: A01207; TIBOC.
DR HSSP: P10646; ITEX.
DR INTERPRO: IPR002223; -.
DR PFAM: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 1.
KW Serine protease inhibitor; Glycoprotein.
FT DISULFID 8 58
FT DISULFID 17 41
FT DISULFID 33 54
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .).
FT ACT_SITE 18 19 REACTIVE BOND (TRYPSIN).
FT ACT_SITE 67 67 E2B2093B7CD207CD CRC64;
SQ SEQUENCE 67 AA; 7511 MW; E2B2093B7CD207CD CRC64;
```

Query Match 23.4%; Score 126; DB 1; Length 67;
Best Local Similarity 41.0%; Pred. No. 8.4e-08;
Matches 25; Conservative 6; Mismatches 26; Indels 4; Gaps 1;

QY 39 DPCKLDNFGSCYEVHFRYFNRTSKRCETFFVSGCNGNLFNFKLIEREVACVAKYKPP 98

Db 6 DLCDLPQARGPCRAALLRYFNSTSNACEPTFYGGCGGNNEF-----ETTEMCLRICEPP 61

QY 99 R 99

Db 62 Q 62

```
RESULT 10
ELAC_TRIVU STANDARD; PRT; 102 AA.
AC Q29143;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE EARLY LACTATION PROTEIN PRECURSOR.
GN ELP.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
```



```

FT CARBOHYD 170 170 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 23 D -> A (IN REF. 4).
SQ SEQUENCE 235 AA; 26934 MW; 975ABA5C53F7C65F CRC64;

Query Match 23.0%; Score 124; DB 1; Length 235;
Best Local Similarity 31.4%; Pred. No. 5.2e-07;
Matches 27; Conservative 18; Mismatches 37; Indels 4; Gaps 1;

QY 18 LNTLL---GGVNKIAEKIGDLKDPCKLDMNFGSCYEVHFRYFVYRTSKRCETVFSG 73
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 9 LSIILLFLTEALGDAQAEFTGNNAEICLLPLDYGPCRALLRYYDRYQSCRFYGG 68
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

QY 74 CNGNLFNFKLKIEREVACVAKYKPPR 99
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 69 CEGNANNFYTWACDADACWRIKVPK 94
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 12
AMB_MOUSE STANDARD; PRT; 349 AA.
AC Q07456; 061294;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AMBP PROTEIN PRECURSOR (CONTAINS: ALPHA-1-MICROGLOBULIN; INTER-ALPHA-
DE TRYPSIN INHIBITOR LIGHT CHAIN (ITI-LC) (BIKUNIN) (HI-30)).
GN AMBP OR ITIL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=LIVER;
RX MEDLINE=93363639; PubMed=7689339;
RA Chan P., Salier J.P.;
RT "Mouse alpha-1-microglobulin/bikunin precursor: cDNA analysis, gene
RT evolution and physical assignment of the gene next to the orosomucoid
RT locus.";
RL Biochim. Biophys. Acta 1174:195-200(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RX MEDLINE=95189774; PubMed=7533761;
RA Itoh H., Ide H., Kataoka H., Tomita M., Yoshihara H., Nawa Y.;
RT "cDNA sequencing of mouse alpha 1-microglobulin/inter-alpha-trypsin
RT inhibitor light chain and its expression in acute inflammation.";
RL J. Biochem. 116:767-772(1994).
RN [3]
RP SEQUENCE OF 128-349 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Itoh H., Ide H., Yoshihara H., Nawa Y.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: ALPHA-1-MICROGLOBULIN OCCURS IN MANY PHYSIOLOGICAL
CC FLUIDS INCLUDING PLASMA, URINE, AND CEREBROSPINAL FLUID. IT
CC APPEARS NOT ONLY AS A FREE MONOMER BUT ALSO IN COMPLEXES WITH IGA
CC AND ALBUMIN (BY SIMILARITY).
CC !- FUNCTION: INTER-ALPHA-TRYPSIN INHIBITOR, PRESENT IN PLASMA AND
CC URINE, INHIBITS TRYPSIN, PLASMIN, AND LYOSOMAL GRANULOCYTIC
CC ELASTASE (BY SIMILARITY).
CC !- SUBUNIT: INTER-ALPHA-TRYPSIN INHIBITOR CONSIST OF A LIGHT CHAIN
CC AND AN HEAVY CHAIN. THERE ARE THREE DIFFERENT HEAVY CHAINS.
CC !- PTM: THE PRECURSOR IS PROTEOLYTICALLY PROCESSED INTO TWO
CC SEPARATELY FUNCTIONING PROTEINS.
CC !- PTM: HC CONTAINS A COVALENTLY LINKED BROWN-YELLOW CHROMOPHORE (BY
CC SIMILARITY).
CC !- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE LIPOCALIN
CC FAMILY.
CC !- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE BPTI/KUNITZ
CC FAMILY OF INHIBITORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X68680; CAA48640.1; -.
CC EMBL: D28812; BAA05973.1; -.
CC HSSP: P12111; IKUN.
CC MGD: MGI:86002; AMBP.
CC INTERPRO: IPR000566; -.
CC INTERPRO: IPR002223; -.
CC INTERPRO: IPR002345; -.
CC PFAM: PF00014; Kunitz_BPTI; 2.
CC PFAM: PF00061; Lipocalin; 1.
CC PRINTS: PR00179; LIPOCALIN.
CC PRINTS: PR00759; BASICPTASE.
CC PROSITE: PS00280; BPTI_KUNITZ_1; 2.
CC PROSITE: PS00279; BPTI_KUNITZ_2; 2.
CC PROSITE: PS00213; LIPOCALIN; 1.
CC Glycoprotein; Plasma; Signal; Serine protease inhibitor; Repeat;
KW Lipocalin.
KW SIGNAL.
FT CHAIN 1 19 BY SIMILARITY.
FT CHAIN 20 202 ALPHA-1-MICROGLOBULIN.
FT CHAIN 205 349 INTER-ALPHA-TRYPSIN INHIBITOR LIGHT
FT CHAIN.
FT DOMAIN 226 281 I.
FT DOMAIN 282 348 II.
FT BINDING 52 52 CHROMOPHORE (BY SIMILARITY).
FT DISULFID 90 187 BY SIMILARITY.
FT DISULFID 230 280 BY SIMILARITY.
FT DISULFID 239 263 BY SIMILARITY.
FT DISULFID 255 276 BY SIMILARITY.
FT DISULFID 286 336 BY SIMILARITY.
FT DISULFID 295 319 BY SIMILARITY.
FT DISULFID 311 332 BY SIMILARITY.
FT CARBOHYD 33 33 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC...) (POTENTIAL).
FT ACT_SITE 240 241 INHIBITORY SITE (PI) (CHYMOTRYPSIN,
FT ELASTASE) (BY SIMILARITY).
FT ACT_SITE 296 297 INHIBITORY SITE (PI) (TRYPSIN) (BY
FT SIMILARITY).
FT CONFLICT 65 65 Q -> S (IN REF. 2).
FT SEQUENCE 349 AA; 39070 MW; CE4D9FC7375DA80B CRC64;

Query Match 23.0%; Score 124; DB 1; Length 349;
Best Local Similarity 49.0%; Pred. No. 7.9e-07;
Matches 25; Conservative 6; Mismatches 18; Indels 2; Gaps 1;

QY 33 ICGDLK--DPCKLDMNFGSCYEVHFRYFVYRTSKRCETVFSGCGNGLNMF 81
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 220 ITGTLKKEDSCQLNYSQPCLGMOERYIYNGASMACEFYGGCLGNGNMF 270
   ||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 13
IVBL_DENPO STANDARD; PRT; 60 AA.
ID IVBL_DENPO
AC P00979;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE VENOM BASIC PROTEASE INHIBITOR I (DENDROTOXIN I).
OS Dendroaspis polylepsis polylepis (Black mamba).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Dendroaspis.
RN [1]
RP SEQUENCE.
RC TISSUE=VENOM;
RX MEDLINE=73174607; PubMed=4512962;
RA Strydom D.J.;
```

[illegible]

FT	DOMAIN	121	171	(VII(A)/TISSUE FACTOR BINDING SITE). BPTI/KUNITZ INHIBITOR 2
FT	DOMAIN	213	263	(FACTOR X(A) BINDING SITE). BPTI/KUNITZ INHIBITOR 3
FT				(BY SIMILARITY).
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				REACTIVE BOND (BY SIMILARITY).
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				REACTIVE BOND (BY SIMILARITY).
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				REACTIVE BOND (BY SIMILARITY).
FT				BY SIMILARITY.
FT				BY SIMILARITY.
FT				REACTIVE BOND (BY SIMILARITY).
FT				N-LINKED (GLCNAC. .) (POTENTIAL).
FT				N-LINKED (GLCNAC. .) (POTENTIAL).
FT				N-LINKED (GLCNAC. .) (POTENTIAL).
FT				MISSING (IN REF. 3).
FT				PKSI -> RNLS (IN REF. 3).
FT				300 AA; 34435 MW; A08DE357708CA6 CRC54;
FT				SEQUENCE

Query Match 22.4%; Score 121; DB 1; Length 300;
Best Local Similarity 47.2%; Pred. No. 1.5e-06;
Matches 25; Conservative 6; Mismatches 22; Indels

QY 39 DPCKLDMFSGCYEVHFRFYFNRTSKRCETVFSGGNGLNLFKLIKIEREVC 91
| | : : | | |||| | : | : | : | : | : | : | : |
pb 119 DFCFLEEDPGICRGYITRYFYNNQKOCERFKYGCGCLGNLNFESLEECKNTC 171

Db 119 DFCLEEDPGICRGYITRYFYNNQSKQCERFKYGGCLGNLNNFESLEECKNTC 171

Search completed: March 24, 2001, 13:24:21
Job time: 167 sec

QY	91 C 91
Db	280 C 280
RESULT	15
TFPI RABIT	
ID	TFPI_RABIT STANDARD; PRT; 300 AA.
AC	R19761; Q28828;
DT	01-FEB-1991 (Rel. 17, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	TISSUE FACTOR PATHWAY INHIBITOR PRECURSOR (TFPI) (LIPOPROTEIN-ASSOCIATED COAGULATION INHIBITOR) (LACI) (EXTRINSIC PATHWAY INHIBITOR) (EPI).
DE	TFPI.
GN	Oryctolagus cuniculus (Rabbit)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OP	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=LIVER:
RA	MEDLINE=91057146; PubMed=2136251;
RX	Wesselschmidt R.L., Girard T.J., Broze G.J. Jr.;
RT	"cDNA sequence of rabbit lipoprotein-associated coagulation inhibitor."
RL	Nucleic Acids Res. 18:6440-6440(1990).
RL	[2]
RP	REVISIONS TO 72: 211 AND 218.
RC	TISSUE=LIVER:
RA	MEDLINE=92335027; PubMed=1630940;
RX	Warn-Cramer B.J., Broze G.J. Jr., Komives E.A.;
RT	"cDNA sequence of rabbit tissue factor pathway inhibitor."
RL	Nucleic Acids Res. 20:3548-3548(1992).
RL	[3]
RP	SEQUENCE FROM N.A.
RC	TISSUE=LUNG:
RA	MEDLINE=93276427; PubMed=8503123;
RX	Belacouaj A., Kuppussawamy M.N., Birktoft J.J., Bajaj S.P.;
RT	"Revised cDNA sequence of rabbit tissue factor pathway inhibitor."
RL	Thromb. Res. 69:547-553(1993).
CC	!- FUNCTION: INHIBITS FACTOR X (X(A)) DIRECTLY AND, IN A XA-DEPENDENT WAY, INHIBITS VII(A)/TISSUE FACTOR ACTIVITY, PRESUMABLY BY FORMING A QUATERNARY X(A)/LACI/VII(A)/TF COMPLEX. IT POSSESSES AN ANTITHROMBOTIC ACTION AND ALSO THE ABILITY TO ASSOCIATE WITH LIPOPROTEINS IN PLASMA.
CC	!- DOMAIN: THIS INHIBITOR CONTAINS THREE INHIBITORY DOMAINS.
CC	!- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
CC	HIGHLY SIMILAR TO TFP2.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annotation or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X54708; CAA38515.1; ALT_SEQ.
DR	EMBL; S61902; AAB26836.1; -
DR	PIR; S12143; S12143.
DR	HSSP; P10646; ITFX.
DR	INTERPRO; IPR002223; -
DR	PFAM; PF00014; Kunitz_BPTI; 3.
DR	PRINTS; PR00759; BASICPTASE.
DR	PROSITE; PS00280; BPTI_KUNITZ_1; 3.
DR	PROSITE; PS0279; BPTI_KUNITZ_2; 3.
KW	Serine protease inhibitor; Glycoprotein; Repeat; Blood coagulation; Signal.
KW	1 24
FT	SIGNAL.
FT	25 300
FT	CHAIN
FT	50 100
FT	DOMAIN
FT	TISSUE FACTOR PATHWAY INHIBITOR.
FT	BPTI/KUNITZ INHIBITOR 1

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:22:25 ; Search time 47.63 Seconds
(without alignments)
71.073 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MKSAKGLRFFIFCSLNT.....NFKLKIERVACVAKPKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	539	100.0	99	20 Y35928	Extended human sec
2	529	98.1	99	20 Y36054	Extended human sec
3	427	79.2	80	20 Y12791	Human 5' EST secre
4	142	26.3	122	20 Y08611	Inter-alpha-trypsi
5	140	26.0	144	18 W25938	SUC2-Ep7-g21-RPDF-
6	138	25.6	144	18 W25936	SUC2-Ep1-g21-RPDF-
7	135	25.0	58	16 R81936	Human ITI-K1 Kunit
8	133.5	24.8	560	15 R62523	Hookworm anticoagu
9	132	24.5	58	16 R78542	Human LACI-K2 deri
10	131.5	24.4	165	18 W25937	Elastase inhibitor
11	131	24.3	58	16 R78554	Human ITI-KuDOM 1
12	131	24.3	95	17 R99213	ITI-D1 Kunitz doma

13	130	24.1	58	17 R99191	Genetically engine
14	129.5	24.0	143	20 Y08609	Inter-alpha-trypsi
15	129.5	24.0	147	17 R92237	Human wild-type ur
16	129.5	24.0	147	18 W25928	Anti-trypsin inhib
17	129.5	24.0	147	19 W69522	rUTI protein seq 1
18	129.5	24.0	352	9 P81110	Sequence of new fu
19	129	23.9	58	17 R99196	Genetically engine
20	128	23.7	58	13 R27402	Human neutrophil e
21	128	23.7	58	17 R99160	Genetically engine
22	127.5	23.7	145	17 R92235	Human urinary tryp
23	127	23.6	55	17 R92217	Inter alpha trypsi
24	127	23.6	58	16 R78553	Human ITI Kunitz d
25	127	23.6	58	17 R99195	Genetically engine
26	127	23.6	58	17 R99197	Genetically engine
27	127	23.6	58	19 W64115	Human Kunitz-type
28	127	23.6	58	20 W92862	US5880256 Seq ID 3
29	127	23.6	111	14 R39807	Synthetic yeast le
30	126	23.4	58	13 R27394	Human neutrophil e
31	126	23.4	58	17 R99157	Human aprotinin-11
32	126	23.4	58	17 R99161	Genetically engine
33	126	23.4	58	17 R99162	Genetically engine
34	126	23.4	124	18 W25933	New protease inhib
35	125	23.2	58	17 R99158	Genetically engine
36	125	23.2	58	17 R99193	Genetically engine
37	125	23.2	58	17 R99164	Genetically engine
38	125	23.2	60	11 R08291	Human bikunin doma
39	125	23.2	74	17 R92238	Human UTI modified
40	125	23.2	124	17 R92233	Human mature urina
41	124	23.0	55	18 W25931	RPDF-Kunitz domain
42	124	23.0	58	17 R99194	Genetically engine
43	124	23.0	124	18 W25932	New protease inhib
44	124	23.0	145	18 W25935	Novel protease inh
45	124	23.0	235	16 R74977	Human Kunitz-type

ALIGNMENTS

RESULT 1
Y35928
ID Y35928 standard; Protein; 99 AA.
XX
XX Y35928;
XX
XX
DF 13-SEP-1999 (first entry)
XX
DE Extended human secreted protein sequence, SEQ ID NO. 177.

Secreted protein; human; cytokine; cellular proliferation; cell movement;
cellular differentiation; immune system regulator; anti-inflammatory;
haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
genetic disease.

OS Homo sapiens.

XX

PN WO9931236-A2.

XX

PD 24-JUN-1999.

XX

PF 17-DEC-1998; 98WO-IB02122.

XX

PR 10-AUG-1998; 98US-0096116.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

XX

PA (GEST) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

DR WPI; 1999-385906/32.

DR N-PSDB; X97612.

XX PT New isolated human secreted proteins
XX PS Claim 9; Page 210; 516pp; English.
XX CC

XX CC This sequence is encoded by an extended human secreted protein coding
XX CC sequence of the invention. The secreted proteins can be used in treating
XX CC or controlling a variety of human conditions. The secreted proteins may
XX CC act as cytokines or may affect cellular proliferation or differentiation
XX CC or may act as immune system regulators, haematopoiesis regulators, tissue
XX CC growth regulators, regulators of reproductive hormones or cell movement
XX CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX CC tumour inhibition activity. The DNAs can be used in forensic procedures
XX CC to identify individuals or in diagnostic procedures to identify
XX CC individuals having genetic diseases resulting from abnormal expression of
XX CC the genes corresponding to the extended cDNAs. They are also useful for
XX CC constructing a high resolution map of the human chromosomes. They can
XX CC also be used for gene therapy to control or treat genetic diseases.

XX SQ Sequence 99 AA;

Query Match 100.0%; Score 539; DB 20; Length 99;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSAKLGFLRLFFIFCSLNTLLGGVKNIAEKICGDLKDPCKLDMNFGSCYEVHFRFYFN 60
Db 1 mksaklgflrlffifcsntlllglggnvkniaekicgdlkdpckldmnmfgscyevhfrfyfn 60
Qy 61 RTSKRCETFFVSCNGNLNFKLIEREVACVAKYKPPR 99
Db 61 rtskrctetfvscngnlnnfnklkierevacvakykppr 99

RESULT 2
Y36054
ID Y36054 standard; Protein; 99 AA.
AC Y36054;
XX 13-SEP-1999 (first entry)
XX DE Extended human secreted protein sequence, SEQ ID NO. 439.

XX KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
XX KW cellular differentiation; immune system regulator; anti-inflammatory;
XX KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
XX KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
XX KW genetic disease.
XX OS Homo sapiens.
XX PN WO9931236-A2.
XX PD 24-JUN-1999.
XX PF 17-DEC-1998; 98WO-IB02122.
XX PR 10-AUG-1998; 98US-0096116.
XX PR 17-DEC-1997; 97US-0069957.
XX PR 09-FEB-1998; 98US-0074121.
XX PR 13-APR-1998; 98US-0081563.
XX PA (GEST) GENSET.
XX PI Bouqueleret L, Duclert A, Dumas Milne Edwards J;
XX DR WPI: 1999-385906/32.
XX DR N-PSDB; X97738.
XX PT New isolated human secreted proteins

PS Claim 9; Page 376; 516pp; English.

XX CC This sequence is encoded by an extended human secreted protein coding
XX CC sequence of the invention. The secreted proteins can be used in treating
XX CC or controlling a variety of human conditions. The secreted proteins may
XX CC act as cytokines or may affect cellular proliferation or differentiation
XX CC or may act as immune system regulators, haematopoiesis regulators, tissue
XX CC growth regulators, regulators of reproductive hormones or cell movement
XX CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX CC tumour inhibition activity. The DNAs can be used in forensic procedures
XX CC to identify individuals or in diagnostic procedures to identify
XX CC individuals having genetic diseases resulting from abnormal expression of
XX CC the genes corresponding to the extended cDNAs. They are also useful for
XX CC constructing a high resolution map of the human chromosomes. They can
XX CC also be used for gene therapy to control or treat genetic diseases.

XX SQ Sequence 99 AA;

Query Match 98.1%; Score 529; DB 20; Length 99;
Best Local Similarity 98.0%; Pred. No. 6.3e-56;
Matches 97; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSAKLGFLRLFFIFCSLNTLLGGVKNIAEKICGDLKDPCKLDMNFGSCYEVHFRFYFN 60
Db 1 mksaklgflrlffifcsntlllglggnvkniaekicgdlkdpckldmnmfgscyevhfrfyfn 60
Qy 61 RTSKRCETFFVSCNGNLNFKLIEREVACVAKYKPPR 99
Db 61 rtskrctetfvscngnlnnfnklkierevacvakykppr 99

RESULT 3
Y12791
ID Y12791 standard; Protein; 80 AA.
AC Y12791;
XX 21-JUN-1999 (first entry)
XX DE Human 5' EST secreted protein SEQ ID NO:381.
XX KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition.
XX OS Homo sapiens.
XX PN WO9906549-A2.
XX PD 11-FEB-1999.
XX PF 31-JUL-1998; 98WO-IB01231.
XX PR 01-AUG-1997; 97US-0905279.
XX PA (GEST) GENSET.
XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX DR WPI: 1999-153779/13.
XX DR N-PSDB; X51569.

XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX PS Claim 34; Page 449; 522pp; English.
XX CC X51459 to X51691 represent 5' expressed sequence tags (ESTs) for human
XX CC secreted proteins, and encode the proteins given in Y12681 to Y12913,

	Misc-difference	54	/note= "mutated amino acid; replaces amino acid Gln in wild type sequence"
	Domain	77..144	/note= "Kunitz domain 2"
	Disulfide-bond	81..131	
	Disulfide-bond	90..114	
	Active-site	91	
		/note= "active site residue"	
	Disulfide-bond	106..127	
	JP09124700-A.		
	13-MAY-1997.		
	07-NOV-1995;	95JP-0288527.	
	07-NOV-1995;	95JP-0288527.	
	(GREC) GREEN CROSS CORP.		
	WPI; 1997-316576/29.		
	N-PSDB; T79086.		
	New protease inhibitor - useful for treating diseases involving elastase		
	Disclosure; Fig 30; 37pp; Japanese.		
	This is the amino acid sequence of the polypeptide encoded by the insert in plasmid pH337 which comprises the novel elastase specific inhibitor Epi-D321-RPDF-52-55 (W25932). The inhibitor sequence is linked downstream of the yeast invertase (SUC2) signal peptide sequence. The modified protease inhibitors are targeted to the protease elastase, especially from neutrophils and can be used to treat diseases associated with elastase. Modifications of the active site were done by site directed mutagenesis.		
	Sequence	144 AA;	
	Query Match	25.6%; Score 138; DB 18; Length 144;	
	Best Local Similarity	41.0%; Pred. No. 4.3e-09;	
	Matches	34; Conservative 9; Mismatches 30; Indels 10; Gaps 2;	
QY	9	LLRFFIFCSLTLGLGVNKTAEIKGDLPCKLDMPFGSCVEVHFHYFYNTSKRCET	58
Ddb	3	llqalf-----llaagaaklsar-----pdfqlgysagpciaffprfyngtsmacqt	52
QY	69	FVFSGCNGNLNNFKLKIEREVAC-	91
Ddb	53	fvyggcmgnfnftekeclqtcs	75
RESULT	7		
R81936	ID	R81936 standard; protein; 58 AA.	
XX	AC	R81936;	
XX	DT	18-MAR-1996 (first entry)	
XX	DE	Human ITI-K1 Kunitz domain derived protein DK1-4.1.1.	
XX	KW	Human ITI-K1; kallikrein; DK1-4.1.1.	
XX	OS	inhibitor; KIP; Kunitz domain; hereditary angioedema.	
XX	PX	Homo sapiens.	
XX	PN	WO9521601-A2.	
XX	PD	17-AUG-1995.	

RESULT	11	
R78554		
ID	R78554	standard; peptide; 58 AA.
XX	R78554;	
XX		
XX	01-WAR-1996	(first entry)
DT		
XX	Human ITI-KuDOM 1	derivative DPI-4.1.1.
DE		
XX		
XX	Human; lipoprotein-associated	coagulation inhibitor; peptide library;
KW	inhibitor; plasmin; bovine;	pancreatic trypsin inhibitor; Kunitz domain;
KW	fibrinolysis; fibrinogenolysis;	bleeding; thrombolytic.
KW		
XX	Synthetic.	
OS		
XX		
Key	Location/Qualifiers	
XX		
FT	Misc-difference 10	
FT	/note=	"residue change: Ser to Glu"
FT		
FT	Misc-difference 15	
FT	/note=	"residue change: Met to Arg"
FT		
FT	Misc-difference 17	
FT	/note=	"residue change: Met to Lys"
FT		
FT	Misc-difference 18	
FT	/note=	"residue change: Thr to Phe"
FT		
FT	Misc-difference 34	
FT	/note=	"residue change: Gln to Val"
FT		
FT	Misc-difference 39	
FT	/note=	"residue change: Met to Gly"
FT		

XX 16-DEC-1994; 94US-0358160.
 XX (PROT-) PROTEIN ENG CORP.
 XX Guterman SK, Kent RB, Ladner RC, Ley AC, Markland W;
 XX Roberts BL;
 XX WPI; 1996-321851/32.
 XX New engineered inhibitors of human neutrophil elastase - contg.
 XX aprotinin-like kunitz domain for treating, e.g. cystic fibrosis or
 XX other respiratory disorders
 XX Claim 1; Page 51; 105pp; English.
 XX Genetically engineered human derived Kunitz domains can be used to
 XX inhibit human neutrophil elastase, an enzyme involved in the
 XX elimination of pathogens and the restructuring of connective tissue.
 XX In cases of reduction of the circulating alpha-1-protease inhibitor
 XX (API or alpha-1antitrypsin), or the inactivation of API by oxidation
 XX (smokers emphysema), extensive destruction of the lung tissue may
 XX result from uncontrolled elastolytic activity of human neutrophil
 XX elastase. Other respiratory disorders such as cystic fibrosis are
 XX thought to be caused by human neutrophil elastase release by
 XX neutrophils. The genetically engineered human derived Kunitz
 XX domains can be used to treat such respiratory disorders. See
 XX R99146-R99211.
 XX Sequence 58 AA;
 XX
 Query Match 24.1%; Score 130; DB 17; Length 58;
 Best Local Similarity 47.2%; Pred. No. 1.4e-08;
 Matches 25; Conservative 8; Mismatches 20; Indels 0; Gaps 0;
 QY 39 DPCKLDNFGSCYEVHRYFYNRYSKRCETVFVSGCNGNLNNFKLKIEREVAC 91
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 Db 3 dcfleesagpcvampfrfyfnnqtkqctcfvfy99cmgngnnfletleecnk 55
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 RESULT 14
 Y08609
 ID Y08609 standard; Protein; 143 AA.
 AC Y08609;
 XX
 XX 05-AUG-1999 (first entry)
 XX Inter-alpha-trypsin inhibitor (BPI type) protein fragment TIHUB1.
 XX APP; beta-amyloid precursor protein; human; transgenic mice; pathology;
 XX Alzheimer's disease; model; therapeutic compound; brain; mechanism;
 XX nerve tissue specific promoter; synthesis; inhibitor; deposition;
 XX plaque formation; treatment.
 XX Unidentified.
 OS
 XX US5912410-A.
 PN
 XX 15-JUN-1999.
 PD
 XX 13-APR-1995; 95US-0422333.
 PF
 XX 21-OCT-1994; 94US-0327381.
 PR 15-JUN-1990; 90US-0538857.
 PR 17-JUN-1991; 91US-0716725.
 PR 13-APR-1995; 95US-0422333.
 XX (SCIO-) SCIOS INC.
 PA
 XX Cordell B;
 PI
 XX

DR WPI; 1999-357231/30.
 XX Transgenic mice useful for studying compounds potentially useful in
 XX the treatment of Alzheimer's disease
 XX Disclosure; Fig 8A; 72pp; English.
 XX This invention describes novel transgenic mice expressing proteins
 XX related to the pathology of Alzheimer's disease and which provide models
 XX for studying potentially therapeutic compounds. The transgenic mice
 XX contain a DNA sequence encoding a beta-amyloid precursor protein (APP)
 XX and a nerve tissue specific promoter operably linked to the beta-APP
 XX allowing its expression to form beta-amyloid protein deposits in the
 XX animal's brain. The transgenic mouse is useful for elucidating the
 XX molecular mechanisms involved in the synthesis of and, more importantly,
 XX inhibiting the synthesis and deposition of beta-amyloid proteins (most
 XX importantly in the brain where plaque formation is associated with
 XX Alzheimer's disease) by inhibiting production and/or increasing cleavage
 XX after production. The transgenic animals provide useful models for
 XX studying the in vivo relationships of the proteins to each other and to
 XX other compounds being tested for their usefulness in treating Alzheimer's
 XX disease.
 XX Sequence 143 AA;
 XX
 Query Match 24.0%; Score 129.5; DB 20; Length 143;
 Best Local Similarity 41.4%; Pred. No. 4.4e-08;
 Matches 29; Conservative 9; Mismatches 25; Indels 7; Gaps 2;
 QY 24 GG--VNKIAEKICGLDKPCIKDMNFGSCYEVHRYFYNRYSKRCETVFVSGCNGNLNNF 81
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 Db 12 ggqlvtvctkk-----edscqlgysagpcmgmtsfyfngtmacetfdggcgmgngnnf 66
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 QY 82 KLIKIEREVAC 91
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 Db 67 vtekeclqtcc 76
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
 RESULT 15
 R92237
 ID R92237 standard; protein; 147 AA.
 AC R92237;
 XX
 XX 27-SEP-1996 (first entry)
 XX Human wild-type urinary trypsin inhibitor.
 XX UTI; kunitz domain; urinary trypsin inhibitor; elastase inhibitor;
 XX recombinant protein production; yeast host cell; Pichia;
 XX site-directed mutagenesis.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX peptide 1..21 /label= signal_peptide
 XX Domain 26..76 /label= kunitz_domain_1
 XX Misc-difference 36..40 /label= pl-P4'
 XX /note= "motif replaced by Ile-Ala-Phe-Phe-Pro
 XX to improve elastase inhibitory activity"
 XX
 XX Domain 78..145 /label= kunitz_domain_2
 XX Disulfide-bond 26..76
 XX Disulfide-bond 35..59
 XX Disulfide-bond 51..72
 XX Disulfide-bond 82..132
 XX Disulfide-bond 91..115
 XX Disulfide-bond 107..128
 XX Modified-site 10

FT. /label= O-linked_glycosylation
FT Modified-site 45
FT /label= N-linked_glycosylation

XX PN WO960303-A1.

XX PD 08-FEB-1996.

XX PF 21-JUL-1995; 95WO-JF01449.

XX PR 21-JUL-1994; 94JP-0169221.

XX PA (GREC) GREEN CROSS CORP.

XX PI Goto T, Horii H, Ideno S;

XX WPI; 1996-117048/12.

XX PT Production of recombinant urinary trypsin inhibitor in Pichia sp. -
PT also Kunitz domain fragments of the inhibitor and new variants
PT having improved elastase inhibitor activity

XX PS Disclosure; Fig 2; 97pp; Japanese.

XX CC Recombinant urinary trypsin inhibitor (rUTI) can be produced in
CC Pichia yeast. By mutating a 5 amino acid motif (P1-P4') within kunitz
CC domain 1 (from MGMTS to IAFPP), the resulting rUTI has improved
CC elastase inhibitory activity. The present sequence is that of
CC wild-type precursor UTI.

XX SQ Sequence 147 AA;

Query Match 24.0%; Score 129.5; DB 17; Length 147;
Best Local Similarity 41.4%; Pred. No. 4.5e-08;
Matches 29; Conservative 9; Mismatches 25; Indels 7; Gaps 2;

QY 24 GG--VNKIAEKIGGDKLPCKLDNMFSGCYEVHFRFYNRYSKRCETFFVSGCGNLNLF 81
|| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

Db 12 ggqlvtevtkk-----edscqlgysagpcmgmtsrlyfngtsmacetfgyggcmgngnnf 66

QY 82 KLIKIEREVAC 91

Db 67 vtekeclqtc 76

Search completed: March 24, 2001, 13:22:26
Job time: 57 sec

THIS PAGE BLANK (USPTO)

[T]
N_N

RESULT 8

[T]
N_N

Query Match 22.3%; Score 120; DB 13; Length 342;
Best Local Similarity 42.4%; Pred. No. 1.6e-06;
Matches 25; Conservative 6; Mismatches 22; Indels 6; Gaps 1;
QY 39 DPKLDMNFGSCYEVHFRFYNYRTSKRCETFFVSGGNGNLNFKLIE-----REVAC 91
DB 223 ESCR LAPSGPCLGNHNRFYNSMACETFTQGGCLGNNNFHSKEKCLHDCRTEAC 281
RESULT 14
Q9NOX7 PRELIMINARY; PRT; 169 AA.
AC Q9NOX7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TROPHOBLAST KUNITZ DOMAIN PROTEIN 2.
GN TKDpl.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BREED SIMMENTAL X HEREFORD;
RA Green J.A., Maclean J.A. II, Gan X., Xie S., Roberts R.M.;
RT "Cloning and characterization of the expression of trophoblast Kunitz
domain proteins (TKDP).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF241777; AAF61248.1; -
SQ SEQUENCE 169 AA; 18855 MW; B65C16F4DF0BA3D CRC64;

Query Match 22.2%; Score 119.5; DB 6; Length 169;
Best Local Similarity 42.9%; Pred. No. 8.6e-07;
Matches 24; Conservative 5; Mismatches 18; Indels 9; Gaps 1;
QY 36 DLKDPCKLDMNFGSCYEVHFRFYNYRTSKRCETFFVSGGNGNLNFKLIEREVAC 91
DB 116 ELKGPCKDQMT-----RYFNKATRYCEPFVYGGCGGNKNFQTLSHCIIVTC 162
RESULT 15
Q922U8 PRELIMINARY; PRT; 246 AA.
AC Q922U8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE TPPIBETA (FRAGMENT).
GN TPPIBETA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Chang J.-Y., Monroe D.M., Roberts H.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF016313; AAD01586.1; -
DR HSP; P10646; ITEX
DR INTERPRO: IPR002223; -
DR PFAM: PF00014; Kunitz_BPTI; 2.
DR PRINTS: PR00759; BASICPTASE.
DR PROSITE: PS00280; BPTI_KUNITZ; 2.
KW Serine protease inhibitor.
FT NON_TER 1
SQ SEQUENCE 246 AA; 28003 MW; 4BEDC23EE04F07E6 CRC64;

Query Match 22.1%; Score 119; DB 11; Length 246;
Best Local Similarity 39.1%; Pred. No. 1.5e-06;
Matches 27; Conservative 8; Mismatches 32; Indels 2; Gaps 1;
QY 25 GVNKIAEKICGDLKDP--CKLDMNFGSCYEVHFRFYNYRTSKRCETFFVSGGNGNLNFK 82
DB 96 GYEXTAVKAASGAERPDFCFLEEDPGLCRGYMKRYLYNNOTKQCERFYVGGCLGNRNFE 155
QY 83 LKIEREVAC 91
DB 156 TLDECKKTC 164
Search completed: March 24, 2001, 13:25:41
Job time: 247 sec

Query Match 22.1%; Score 119; DB 11; Length 246;
Best Local Similarity 39.1%; Pred. No. 1.5e-06;
Matches 27; Conservative 8; Mismatches 32; Indels 2; Gaps 1;
QY 25 GVNKIAEKICGDLKDP--CKLDMNFGSCYEVHFRFYNYRTSKRCETFFVSGGNGNLNFK 82
DB 96 GYEXTAVKAASGAERPDFCFLEEDPGLCRGYMKRYLYNNOTKQCERFYVGGCLGNRNFE 155
QY 83 LKIEREVAC 91
DB 156 TLDECKKTC 164
Search completed: March 24, 2001, 13:25:41
Job time: 247 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:12 ; Search time 44.26 Seconds
(without alignments)
151.879 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MSAKLGFLRRFFIFCSLNT.....NFKLKIEREVACVAKYKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	28.6	110	1 TITROR	basic proteinase i
2	134	24.9	352	1 TIBOBI	alpha-1-microglobu
3	133	24.7	1558	2 T34394	hypothetical prote
4	130	24.1	62	2 S07451	proteinase inhibit
5	129.5	24.0	352	1 HCHU	alpha-1-microglobu
6	128	23.7	2167	2 T34395	hypothetical prote
7	127	23.6	65	1 TIVIVC	venom basic protei
8	127	23.6	123	2 A29652	inter-alpha-trypsi
9	127	23.6	125	1 TIBOBI	alpha-1-microglobu
10	126	23.4	67	1 TIBOC	trypsin inhibitor,
11	126	23.4	102	2 S69288	early lactation pr
12	124.5	23.1	249	2 T32060	hypothetical prote
13	124	23.0	235	2 A54951	tissue factor path
14	124	23.0	349	2 S35708	alpha-1-microglobu
15	123.5	22.9	2225	2 T26063	hypothetical prote
16	122	22.6	60	1 TIEPVI	venom basic protei
17	122	22.6	349	2 S21089	alpha-1-microglobu
18	121	22.4	299	2 I46937	tissue factor path
19	121	22.4	300	2 S12143	lipoprotein-associ
20	120	22.3	304	1 JC2264	tissue factor path
21	118.5	22.0	396	2 S53325	tissue factor path
22	118	21.9	57	2 A59204	basic proteinase i
23	118	21.9	62	2 S19327	venom basic protei
24	118	21.9	337	1 TIFGBI	alpha-1-microglobu
25	117	21.7	57	1 TIFHBP	proteinase inhibit
26	117	21.7	302	1 TIRTKG	tissue factor path
27	117	21.7	805	2 T34212	hypothetical prote
28	116	21.5	304	1 TIHUGK	tissue factor path
29	116	21.5	1743	2 T26859	hypothetical prote

30	115.5	21.4	1965	2 T33216	hypothetical prote
31	115	21.3	59	1 VIEPIA	venom basic protei
32	114	21.2	62	2 A4180	taicatoxin serine
33	113	21.0	1599	2 T16210	hypothetical prote
34	112.5	20.9	355	1 S22181	gamma-1-microglobu
35	112	20.8	1043	2 T19734	hypothetical prote
36	111	20.6	57	2 S12957	venom animal Kunit
37	111	20.6	372	2 JC2556	alpha-1-microglobu
38	110	20.4	252	2 JG0185	hepatocyte growth
39	110	20.4	922	2 T23573	hypothetical prote
40	109	20.2	60	2 A36989	calcicudine - eas
41	109	20.2	2944	2 A54849	collagen alpha 1(V
42	108.5	20.1	765	2 S42880	anyloid precursor-
43	108	20.0	62	2 S01803	chymotrypsin inhib
44	107.5	19.9	751	2 A49574	beta-amyloid precu
45	107.5	19.9	763	2 A49321	anyloid beta (A4)

ALIGNMENTS

RESULT 1

TITROR

basic proteinase inhibitor - loggerhead

C:Species: Caretta caretta (loggerhead)

C:Date: 31-May-1979 #sequence_revision 23-Oct-1981 #text_change 05-Aug-1994

C:Accession: A01224

R:Kato, I.; Tominaga, N.

Fed. Proc. 38, 832, 1979

A:Title: Trypsin--subtilisin inhibitor from red sea turtle eggwhite consists of two t

A:Reference number: A01224

A:Accession: A01224

A:Molecule type: protein

A:Residues: 1-110 <KAT>

C:Comment: This inhibitor, isolated from egg white, consists of two nonhomologous dom

C:Superfamily: loggerhead basic proteinase inhibitor; animal Kunitz-type proteinase i

C:Keywords: pyroglutamic acid; serine proteinase inhibitor

F:8-58/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:63-105/Domain: antileukoproteinase repeat homology <ALP>

F:/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:8-58,17-41,33-54,76-97,80-92,86-101/Disulfide bonds: #status predicted

F:18/Inhibitory site: Lys (trypsin) #status predicted

Query Match 28.6%; Score 154; DB 1; Length 110;

Best Local Similarity 46.9%; Pred. No. 5.4e-10;

Matches 30; Conservative 7; Mismatches 23; Indels 4; Gaps 1;

QY 35 GDLKDPCKLDNFGSCYEVHFRFYFNRTSKRCETVFSGCGNGLNFKLKIEREVACVAK 94

|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

Db 2 GDRKDICRLPPEQGPCGKGRIPRYFNYPASRMCSFIYGGCKGNKNTKAE---CVRA 57

QY 95 YKPP 98

||

Db 58 CRPP 61

RESULT 2

TIBOBI

alpha-1-microglobulin / inter-alpha-trypsin inhibitor precursor [validated] - bovine

N:Alternate names: BT-14 (inhibitory fragment of ITI); bikunin; ITI

C:Species: Bos primigenius taurus (cattle)

C:Date: 25-Feb-1985 #sequence_revision 04-Feb-2000 #text_change 18-Aug-2000

C:Accession: S68149; A91717; A90685; S31219; A01209

R:Lindqvist, A.; Akerstrom, B.

Biochim. Biophys. Acta 1306, 98-106, 1996

A:Title: Bovine alpha(1)-microglobulin/bikunin. Isolation and characterization of liv

A:Reference number: S68149; MUID:96201710

A:Accession: S68149

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-352 <LIN>

A:Cross-references: EMBL:U35642; NID:g1016297; PIDN:AAB07599.1; PID:g1016298

```

Db      1497  CQHKNACQLPKVQGPCSGKHSHSYYYNTASHOCETFTYGGCLGNTNRNRTATIECQARCPSS 1556
          QY           94 KY 95
          :         :
Db      1557 KF 1558
          :
RESULT   4
S07451
Proteinase inhibitor 5.II - snake-locks sea anemone
C:Species: Anemonia sulcata (snake-locks sea anemone)
C>Date: 31-Dec-1990 #sequence_revision 09-May-1997 #text_change 09-May-1997
C:Accession: S07451; B27222
R:Wunderer, G.; Machleidt, W.; Fritz, H.
Meth. Enzymol. 80, 816-820, 1981
A:Title: The broad-specificity proteinase inhibitor 5 II from the sea anemone Anemonia
A:Reference number: S07451
A:Accession: S07451
A:Molecule type: protein
A:Residues: 1-59 <WN>
A>Note: 13-Arg, 16-Gly, 17-Gly, 25-Leu, 28-Arg, and 39-Arg were also found
R:Krebs, H.C.; Habermehl, G.G.
Naturwissenschaften 74, 395-396, 1987
A:Title: Isolierung und strukturaufklaerung eines haemolytisch aktiven peptids aus de
A:Reference number: A94700
A:Accession: B27222
A:Molecule type: protein
A:Residues: 1-38, 'R', '40', 'B', '42', 'BB', '45-48', 'ZZ', '51', 'Z', '53-62 <KRE>
C:Superfamily: basic proteinase inhibitor; animal kunitz-type proteinase inhibitor ho
C:Keywords: serine proteinase inhibitor
F:5-55/Domain: animal kunitz-type proteinase inhibitor homology <BPI>

Query Match      24.1%; Score 130; DB 2; Length 62;
Best Local Similarity 45.8%; Pred. No. 1.4e-07;
Matches 27; Conservative 5; Mismatches 23; Indels 4; Gaps 1;

QY    33  ICGDLKPDKLDNMFNGSCYEYHFNFYFNRTSKRCETFVFGCCGNLNNFKLTIEREVAC 91
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     1  INGD----CELPKVVGPCRRAFPRYYNTSSKKRCETFIYGGCGGNANFNFTLECEKVC 55

```

RESULT 5
HCHU
alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor - human
N:Alternate names: bikunin; complex-forming glycoprotein heterogeneous in charge (HC)
rich protein
N:Contains: alpha-1-microglobulin (protein HC); inter-alpha-trypsin inhibitor
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1982 #sequence_revision 30-Jun-1987 #text_change 04-Feb-2000
C:Accession: S13433; S10778; A93642; A90074; A90225; A90686; PNO450; B39079; A61580;
3217
R:Vetr, H.; Gebhard, W.
Biol. Chem. Hoppe-Seyler 371, 1185-1196, 1990
A:Title: Structure of the human alpha(1)-microglobulin-bikunin gene.
A:Reference number: S13433; MUID: 91214554
A:Accession: S13433
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <VET1>
A:Cross-references: EMBL:X54816; NID:q24475; PIDN:CAA38585.1; PID:q825614; EMBL:X54811
R:Diarra-Mehrpour, M.; Bourguignon, J.; Sesboue, R.; Sallier, J.P.; Leveillard, T.; M
Eur. J. Biochem. 191, 131-139, 1990
A:Title: Structural analysis of the human inter-alpha-trypsin inhibitor light-chain 9
A:Reference number: S10778; MUID: 90336621
A:Accession: S10778
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <DIA>
R:Kaumeyer, J.F.; Polazzi, J.O.; Kotick, M.P.
Nucleic Acids Res. 14, 7839-7850, 1986
A:Title: The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alpha-

A:Reference number: A93642; MUID:87040757
 A:Accession: A93642
 A:Molecule type: mRNA
 A:Residues: 1-352 <KAD>
 A:Cross-references: GB:X04494; NID:g24478; PIDN:CAA28182.1; PID:g24479
 R:Lopez Otin, C.; Grubb, A.O.; Mendez, E.
 Arch. Biochem. Biophys. 228, 544-554, 1984
 A:Title: The complete amino acid sequence of human complex-forming glycoprotein heteroglycin
 A:Reference number: A90074; MUID:84126849
 A:Accession: A90074
 A:Molecule type: protein
 A:Residues: 20-56,58-202 <LOP>
 A:Experimental source: individual with tubular proteinuria
 A:Note: no evidence of sequence heterogeneity could be found, in spite of persistent heterogeneity
 R:Takagi, T.; Takagi, K.; Kawai, T.
 Biochem. Biophys. Res. Commun. 98, 997-1001, 1981
 A:Title: Complete amino acid sequence of human alpha-1-microglobulin.
 A:Reference number: A90225; MUID:81184038
 A:Accession: A90225
 A:Molecule type: protein
 A:Residues: 20-47;58-136,138-141,'T',143-144,146-198 <TAK>
 A:Experimental source: pooled urine of patients with tubular proteinuria
 R:Reisinger, P.; Hochstrasser, K.; Albrecht, G.J.; Lempart, K.; Salier, J.P.
 Biol. Chem. Hoppe-Seyler 366, 479-483, 1985
 A:Title: Human inter-alpha-trypsin inhibitor: localization of the kunitz-type domains in human plasma
 A:Reference number: A90686; MUID:85225968
 A:Accession: A90686
 A:Molecule type: protein
 A:Residues: 206-290, VI',293-342,'E',344-350 <REI>
 R:Atmani, F.; Lacour, B.; Strecker, G.; Parvy, P.; Druceke, T.; Daudon, M.
 Biochem. Biophys. Res. Commun. 191, 1158-1165, 1993
 A:Title: Molecular characteristics of uronic-acid-rich protein, a strong inhibitor of chondroitinase ABC
 A:Reference number: PN0450; MUID:93221481
 A:Accession: PN0450
 A:Molecule type: protein
 A:Residues: 206-214,'X' <ATM1>
 R:Engchild, J.J.; Salvesen, G.; Hefta, S.A.; Thøgersen, I.B.; Rutherford, S.; Pizzo, S.V.
 J. Biol. Chem. 266, 747-751, 1991
 A:Title: Chondroitin 4-sulfate covalently cross-links the chains of the human blood protein chondroitin sulfate proteoglycan
 A:Reference number: A39079; MUID:91093267
 A:Accession: B39079
 A:Molecule type: protein
 A:Residues: 206-225 <ENG1>
 R:Chirat, F.; Balduyck, M.; Mizon, C.; Laroui, S.; Sautiere, P.; Mizon, J.
 Int. J. Biochem. 23, 1201-1203, 1991
 A:Title: A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inhibitor
 A:Reference number: A61580; MUID:92175157
 A:Accession: A61580
 A:Molecule type: protein
 A:Residues: 214,'X',216-222,'X' <CHI>
 R:McKeehan, W.L.; Sakagami, Y.; Hoshi, H.; McKeehan, K.A.
 J. Biol. Chem. 261, 5378-5383, 1986
 A:Title: Two apparent human endothelial cell growth factors from human hepatoma cells and their cDNAs
 A:Reference number: A92583; MUID:86168278
 A:Accession: B25604
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230,'X',232-239,'X',241-248,'XX',251-252,'X',254 <MCK>
 R:Engchild, J.J.; Thøgersen, I.B.; Pizzo, S.V.; Salvesen, G.
 J. Biol. Chem. 264, 15975-15981, 1989
 A:Title: Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-alpha-1-trypsin inhibitor
 A:Reference number: A92736; MUID:89380192
 A:Accession: C34245
 A:Molecule type: protein
 A:Residues: 206-225 <ENG2>
 R:Traboni, C.; Cortese, R.
 Nucleic Acids Res. 14, 6340, 1986
 A:Title: Sequence of a full length cDNA coding for human protein HC (alpha-1-microglobulin)
 A:Reference number: A25303; MUID:86312901
 A:Accession: A25303
 A:Molecule type: mRNA
 A:Residues: 1-218,'HW' <TRA>
 A:Note: this mRNA sequence appears to contain errors after residue 218
 R:Calero, M.; Escribano, J.; Grubb, A.; Mendez, E.
 J. Biol. Chem. 269, 384-389, 1994
 A:Title: Location of a novel type of interpolypeptide chain linkage in the human protein HC
 A:Reference number: A53110; MUID:94103241
 A:Accession: A53110
 A:Molecule type: protein
 A:Residues: 45-57 <CAL1>
 R:Vetr, H.; Koegler, M.; Gebhard, W.
 FEBS Lett. 245, 137-140, 1989
 A:Title: The domain structure of the inhibitor subunit of human inter-alpha-trypsin I
 A:Reference number: S03552; MUID:89171290
 A:Accession: S03552
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 206-352 <VET2>
 R:Walki, N.; Balduyck, M.; Mases, P.; Capon, C.; Mizon, C.; Han, K.K.; Tartar, A.; Fou
 Biol. Chem. Hoppe-Seyler 373, 1009-1018, 1992
 A:Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolat
 A:Reference number: S28928; MUID:93039735
 A:Accession: S28930
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-215 <MAL>
 R:Morelle, W.; Capon, C.; Balduyck, M.; Sautiere, P.; Kouach, M.; Michalski, C.; Four
 Eur. J. Biochem. 221, 881-888, 1994
 A:Title: Chondroitin sulphate covalently cross-links the three polypeptide chains of
 A:Reference number: S43466; MUID:94229087
 A:Accession: S43466
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-221 <MOR>
 R:Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilcek, J.
 Biochemistry 33, 7423-7429, 1994
 A:Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable co
 A:Reference number: A53642; MUID:94271799
 A:Accession: A53642
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 206-217 <WIS>
 R:Calero, M.; Mendez, E.; Garcia, E.
 Biochim. Biophys. Acta 1249, 91-99, 1995
 A:Title: Expression of the human complex-forming glycoprotein HC (alpha-1-microglobul
 A:Reference number: S55688; MUID:95284116
 A:Accession: S55688
 A:Molecule type: protein
 A:Residues: 20-24 <CAL2>
 R:Bourguignon, J.; Diarra-Mehrpour, M.; Sesboue, R.; Frain, M.; Sala-Trepat, J.M.; Ma
 Biochem. Biophys. Res. Commun. 131, 1146-1153, 1985
 A:Title: Human inter-alpha-trypsin-inhibitor: characterization and partial nucleotide
 A:Reference number: I52208; MUID:86025577
 A:Accession: I52208
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 302-352 <BOU>
 A:Cross-references: GB:M11562; NID:g186587; PIDN:AAA59194.1; PID:g307077
 R:Wojsik, E.G.C.; van den Berg, M.; van der Linden, I.K.; Poort, S.R.; Cupers, R.; Be
 Biochem. J. 311, 753-759, 1995
 A:Title: Factor IX Zuthphen: a Cys(18) -> Arg mutation results in formation of a heter
 A:Reference number: S59509; MUID:96067589
 A:Accession: S59509
 A:Molecule type: protein
 A:Residues: 27-35,'Y',37 <WOJ>
 R:Atmani, F.; Mizon, J.; Khan, S.R.
 Eur. J. Biochem. 236, 984-990, 1996
 A:Title: Identification of uronic-acid-rich protein as urinary bikunin, the light cha
 A:Reference number: S66434; MUID:96270753
 A:Accession: S66434
 A:Molecule type: protein
 A:Residues: 206-214,'X',216-230 <ATM2>
 R:Akerstroem, B.; Bratt, T.; Engchild, J.J.
 FEBS Lett. 362, 50-54, 1995
 A:Title: Formation of the alpha(1)-microglobulin chromophore in mammalian and insect
 A:Reference number: S68728; MUID:95212582
 A:Accession: S68728

```

Query Match      23.7%; Score 128; DB 2; Length 2167;
Best Local Similarity 41.8%; pred. No. 7.6e-06;
Matches 23; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

Qy 37 LKDCKLDNFGSCVEVHFYFYNTSKRCETFFVFGCGNGLNNFKLIERVAC 91
    :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Query Match      23.7%; Score 128; DB 2; Length 2167;
Best Local Similarity 41.8%; pred. No. 7.6e-06;
Matches 23; Conservative 9; Mismatches 23; Indels 0; Gaps 0;

Qy 37 LKDCKLDWDFGSCVEVHFYFYNTSKRCETFFVFGCGNGLNNFKLIEREVC 91
    :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

TTH01

alpha-1-microglobulin/inter-alpha-trypsin inhibitor - horse (fragment)

N:Alternate names: EI-14 (inhibitory fragment of III); III; trypsin inhibitor, E-UTI

C:Species: Equus caballus (domestic horse)

C:Date: 30-Jun-1987 #sequence_revision 04-Feb-2000 #text_change 05-May-2000

C:Accession: A01210; A45653

R:Hochstrasser, K.; Wächter, E.; Albrecht, G.J.; Reisinger, P.

Blot. Chem. Hoppe-Seyler 366, 473-478, 1985

A:Title: Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter-

A:Reference number: A90685; MUID:85225967

```

Qy 39 DPCKLDNFGSCYEVHFRFYFNRTSKRCETVFVSGCNGNLNFKLKIEREVACVAKYKPP 98
      I : I | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 DLCLPQARGPCKAALLRFFYNSTSNACEPTYGGCQGNNEF-----ETTEMCLRICEPP 61
      QY 99 R 99
      :
Db 62 Q 62

RESULT 11
S69288
early lactation protein precursor - brush-tailed possum
C:Species: Trichosurus vulpecula (brush-tailed possum)
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C:Accession: S69288; S74274
R:Piotte, C.P.; Grigor, M.R.
Arch. Biochem. Biophys. 330, 59-64, 1996
A:Title: A novel marsupial protein expressed by the mammary gland only during the ear
A:Reference number: S69288; MUID:96230240
A:Accession: S69288
A:Molecule type: mRNA
A:Residues: 1-102 <Pio>
A:Cross-references: EMBL:U34208; NID:g1565292; PIDN:AAB08977.1; PID:g1002806
A:Accession: S74274
A:Molecule type: protein
A:Residues: 21-30 <Pit>
C:Superfamily: animal Kunitz-type proteinase inhibitor homology
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-102/Product: early lactation protein #status experimental <MAT>
F:43-93/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match 23.4%; Score 126; DB 2; Length 102;
Best Local Similarity 56.1%; Pred. No. 6.1e-07;
Matches 23; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

QY 41 CKLDNFGSCYEVHFRFYFNRTSKRCETVFVSGCNGNLNFF 81
      I : I | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 CLLPSGRGNCOSQILRYFNATSHTCVFVLYSGCNGNGNFF 83
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
T32060
hypothetical protein R12A1.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32060
R:Pauley, A.; Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid R12A1.
A:Reference number: Z21118
A:Accession: T32060
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-249 <PAU>
A:Cross-references: EMBL:AF016680; PIDN:AAB66164.1; GSPDB:GN00023; CESP:R12A1.3
A:Experimental source: strain Bristol N2; clone R12A1
C:Genetics:
A:Gene: CESP:R12A1.3
A:Map position: 5
A:Introns: 75/1; 139/1

Query Match 23.1%; Score 124.5; DB 2; Length 249;
Best Local Similarity 38.5%; Pred. No. 2.2e-06;
Matches 25; Conservative 10; Mismatches 19; Indels 11; Gaps 2;

QY 26 VNKAETKCGDLKDP-----CKLDNFGSCYEVHFRFYFNRTSKRCETVFVSGCNGN 77
      I : I | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 123 VTNSATRIC-----RDPVGIASTISGLPLAVGSCAPAVFYIDASSGRCNQFMYSGCCGN 179
      QY 78 LNNEK 82

```

Db 180 ANNFQ 184

|||||

RESULT 13

A54951

N:Alternate names: placental inhibitor-2 precursor - human

C:Species: Homo sapiens (man)

C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 16-Jul-1999

C:Accession: A54951; I55185; A34029; C34029; B34029

R:Sprecher, C.A.; Kisiel, W.; Mathewes, S.; Foster, D.C.

Proc. Natl. Acad. Sci. U.S.A. 91, 3353-3357, 1994

A:Title: Molecular cloning, expression, and partial characterization of a second human t

A:Reference number: A54951; MUID:94211862

A:Accession: A54951

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-235 <RES>

A:CROSS-references: GB:L27624; NID:9441149; PIDN:AAA20094.1; PID:9441150

A:Experimental source: placenta

R:Miyagi, Y.; Koshikawa, N.; Yasumitsu, H.; Miyagi, E.; Hirahara, F.; Aoki, I.; Misugi, J. Biochem. 116, 939-942, 1994

A:Title: cDNA cloning and mRNA expression of a serine proteinase inhibitor secreted by c

A:Reference number: I55185; MUID:95204397

A:Accession: I55185

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: mRNA

A:Residues: 1-235 <RES>

A:CROSS-references: GB:D29992; NID:9484050; PIDN:BAA06272.1; PID:9484051

A:Note: parts of this sequence, including the amino end of the mature protein, were dete

R:Buetzow, R.; Huhtala, M.L.; Bohn, H.; Virtanen, I.; Seppaelae, M.

Biochem. Biophys. Res. Commun. 150, 483-490, 1988

A:Title: Purification and characterization of placental protein 5.

A:Reference number: A34029; MUID:88106628

A:Accession: A34029

A:Molecule type: protein

A:Residues: 'A', 24-33, 'X', 35 <BU2>

A:Accession: C34029

A:Molecule type: protein

A:Residues: 47-50, 'X', 52-53 <BU2>

A:Accession: B34029

A:Molecule type: protein

A:Residues: 133, 'X', 135-137, 'X', 139-140, 'X', 142-144, 'X', 146 <BU3>

C:Genetics:

A:Gene: GDB:TFPI2

A:CROSS-references: GDB:354485

C:Superfamily: tissue factor pathway inhibitor; animal Kunitz-type proteinase inhibitor

C:Keywords: anticoagulant; blood coagulation; duplication; glycoprotein; heparin binding

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-235/Product: tissue factor pathway inhibitor-2 #status predicted <MAT>

F:36-86/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:96-149/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

F:158-208/Domain: animal Kunitz-type proteinase inhibitor homology <BP3>

F:36-86, 45-69, 61-82, 96-149, 106-130, 122-145, 158-208, 167-191, 183-204/Disulfide bonds: #sta

Query Match 23.0%; Score 124; DB 2; Length 235;

Best Local Similarity 31.4%; Pred. No. 2.3e-06;

Matches 27; Conservative 18; Mismatches 37; Indels 4; Gaps 1;

QY 18 LNTLLG----GGVNKIAETGCKLDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSG 73

|||||

Db 9 LSULLFLTEALGDAQEPPTGNAETCLLPIDYGPCRALLRLYYDYRTQSCRQFLYGG 68

QY 74 CNGNLFNFKLIEREVACVAKYKPPR 99

|||||

Db 69 CEGNANNFYTWACDDACWRIKVPK 94

RESULT 14

S35708

alpha-1-microglobulin / bikunin precursor - mouse

N:Alternate names: alpha 1-microglobulin / inter-alpha-trypsin inhibitor light chain

C:Species: Mus musculus (house mouse)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Jun-2000

C:Accession: S35708; JX0355

R:Chan, P.; Salier, J.P.

Biochim. Biophys. Acta 1174, 195-200, 1993

A:Title: Mouse alpha-1-microglobulin/bikunin precursor: cDNA analysis, gene evolution

A:Reference number: S35708; MUID:93363639

A:Accession: S35708

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-349 <CHA>

A:CROSS-references: EMBL:X68680; NID:9311702; PIDN:CAA48640.1; PID:9311703

R:Itoh, H.; Ide, H.; Kataoka, H.; Tomita, M.; Yoshihara, H.; Nawa, Y.

J. Biochem. 116, 767-772, 1994

A:Title: cDNA sequencing of mouse alpha1-microglobulin/inter-alpha-trypsin inhibitor

A:Reference number: JX0355; MUID:95189774

A:Accession: JX0355

A:Molecule type: mRNA

A:Residues: 1-64, 'S', 66-349 <ITO>

A:CROSS-references: DDBJ:D28812; NID:9556530; PIDN:BAA05973.1; PID:9556531

A:Experimental source: liver

C:Superfamily: protein HC; animal Kunitz-type proteinase inhibitor homology; lipocall

C:Keywords: glycoprotein; inflammation; serine proteinase inhibitor

F:34-187/Domain: lipocalin homology <LIP>

F:230-280/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

F:286-336/Domain: animal Kunitz-type proteinase inhibitor homology <BP2>

F:114, 233/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.0%; Score 124; DB 2; Length 349;

Best Local Similarity 49.0%; Pred. No. 3.4e-06;

Matches 25; Conservative 6; Mismatches 18; Indels 2; Gaps 1;

QY 33 ICGDLK--DPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSGCGNLLNF 81

|||||

Db 220 ITGTLKEDSCOLYSEGPLGMQERYIYNGASMACETFFYGGCLGNGNFF 270

RESULT 15

T26063

hypothetical protein W01F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26063

R:Cummings, P.

submitted to the EMBL Data Library, March 1997

A:Reference number: T20145

A:Accession: T26063

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2225 <WIL>

A:CROSS-references: EMBL:292815; PIDN:CAB07294.1; GSPDB:GN00023; CESP:W01F3.3

A:Experimental source: clone W01F3

C:Genetics:

A:Gene: CESP:W01F3.3

A:Map position: 5

A:Introns: 33/1; 56/1; 100/1; 142/3; 271/3; 451/1; 525/3; 774/1; 1093/1; 1178/1; 1221

Query Match 22.9%; Score 123.5; DB 2; Length 2225;

Best Local Similarity 35.4%; Pred. No. 2.4e-05;

Matches 28; Conservative 12; Mismatches 32; Indels 7; Gaps 2;

QY 19 NTLILGGVNKIAETGCKLDPCKLDMNFGSCYEVHFRFYFNRTSKRCETFFVFSGCGNGL 78

|||||

Db 759 NTVLLGGI-----EDTTDSVNRCLHPRDSGNCRGQVRFWFFDEKKKNDVFYTYCCGNG 814

QY 79 NNFKLIEREVACVAKYKPP 97

|||||

Db 815 NNFASKECMCAIC---HKP 830

Search completed: March 24, 2001, 13:23:14
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:23:51 ; Search time 34.59 Seconds
(without alignments)
51.395 Million cell updates/sec

Title: US-09-215-435-177

Perfect score: 539

Sequence: 1 MSAKLGFLRRFFIFCSLNT.....NFKLKIERVACVAKYKPPR 99

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6_COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	28.6	66	1 US-08-358-160-113	Sequence 113, App
2	154	28.6	79	5 5466783-7	Patent No. 5466783
3	142	26.3	60	1 US-08-358-160-153	Sequence 153, App
4	142	26.3	122	2 US-08-422-333-12	Sequence 12, App
5	142	26.3	122	5 5187153-20	Patent No. 5187153
6	142	26.3	122	5 5220013-23	Patent No. 5220013
7	135	25.0	58	1 US-08-676-125A-46	Sequence 46, App
8	135	25.0	58	2 US-09-136-012A-46	Sequence 46, App
9	132	24.5	58	3 US-08-676-124-65	Sequence 65, App
10	132	24.5	58	3 US-09-414-878-65	Sequence 65, App
11	132	24.5	58	3 US-09-240-136-65	Sequence 65, App
12	131	24.3	58	3 US-08-676-124-77	Sequence 77, App
13	131	24.3	58	3 US-09-414-878-77	Sequence 77, App
14	131	24.3	58	3 US-09-240-136-77	Sequence 77, App
15	131	24.3	484	1 US-08-358-160-7	Sequence 7, App
16	130	24.1	62	1 US-08-358-160-97	Sequence 97, App
17	129.5	24.0	143	2 US-08-422-333-10	Sequence 10, App
18	129.5	24.0	143	5 5223482-20	Patent No. 5223482
19	129.5	24.0	144	5 5187153-18	Patent No. 5187153
20	129.5	24.0	147	1 US-08-358-160-72	Sequence 72, App
21	129	23.9	56	1 US-08-358-160-87	Sequence 87, App
22	129	23.9	67	1 US-08-358-160-123	Sequence 123, App
23	128.5	23.8	122	5 5466783-23	Patent No. 5466783
24	128	23.7	58	1 US-08-358-160-17	Sequence 17, App
25	127	23.6	58	1 US-08-384-489-10	Sequence 10, App
26	127	23.6	58	1 US-08-358-160-86	Sequence 86, App
27	127	23.6	58	1 US-08-463-155A-38	Sequence 38, App
28	127	23.6	58	1 US-08-463-432B-38	Sequence 38, App

ALIGNMENTS

RESULT 1

US-08-358-160-113

; Sequence 113, Application US/08358160

; Patent No. 5663143

; GENERAL INFORMATION:

; APPLICANT: LEY, Arthur C.

; APPLICANT: LADNER, Robert C.

; APPLICANT: GUTERMAN, Sonia K.

; APPLICANT: ROBERTS, Bruce L.

; APPLICANT: MARKLAND, William

; APPLICANT: KENT, Rachel B.

; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ

; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE

; NUMBER OF SEQUENCES: 234

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W. Suite 300

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/358,160

; FILING DATE: 16-DEC-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/133,031

; FILING DATE: 13-OCT-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/009,319

; FILING DATE: 26-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/664,989

; FILING DATE: 01-MAR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/487,063

; FILING DATE: 02-MAR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/240,160

; FILING DATE: 02-SEP-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Cooper, Iver P.

; REGISTRATION NUMBER: 28,005

; REFERENCE/DOCKET NUMBER: LEY-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

Sequence 45, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 38, Appl
Sequence 45, Appl
Sequence 76, Appl
Sequence 76, Appl
Sequence 76, Appl
Sequence 128, App
Sequence 132, App
Sequence 92, Appl
Sequence 18, Appl
Patent No. 5466783
Patent No. 5466783
Sequence 10, Appl
Sequence 10, Appl

29 127 23.6 58 1 US-08-676-125A-45
30 127 23.6 58 1 US-08-206-310A-38
31 127 23.6 58 2 US-08-398-010A-38
32 127 23.6 58 2 US-08-398-628A-38
33 127 23.6 58 2 US-08-399-115A-38
34 127 23.6 58 2 US-09-136-012A-45
35 127 23.6 58 3 US-08-676-124-76
36 127 23.6 58 3 US-09-414-878-76
37 127 23.6 58 3 US-09-240-136-76
38 127 23.6 60 1 US-08-358-160-128
39 127 23.6 60 1 US-08-358-160-132
40 127 23.6 65 1 US-08-358-160-92
41 127 23.6 111 1 US-08-321-658B-18
42 127 23.6 123 5 5466783-21
43 127 23.6 127 5 5466783-24
44 126 23.4 56 1 US-07-700-526-10
45 126 23.4 56 4 PCT-US92-03132-10


```

; Sequence 65, Application US/09240136
; Patent No. 6103499
; GENERAL INFORMATION:
; APPLICANT: DYAX CORP
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert C
; TITLE OF INVENTION: Inhibitors of Human Plamin Derived
; TITLE OF INVENTION: From The Kunitz Domains
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yankwich & Associates
; STREET: 130 Bishop Allen Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5-inch diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Microsoft Windows 98
; SOFTWARE: Microsoft Word 97 SR-1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/240,136
; FILING DATE: (concurrently herewith)
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/676,124
; FILING DATE: 07-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00298
; FILING DATE: 11-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/208,265
; FILING DATE: 10-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/179,685
; FILING DATE: 11-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: YANKWICH, Leon R
; REGISTRATION NUMBER: 30,237
; NAME: ZWICKER, Kenneth P
; REGISTRATION NUMBER: 43,310
; REFERENCE/DOCKET NUMBER: DYX-007.2P US-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-491-4343
; TELEFAX: 617-491-8801
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-240-136-65

Query Match 24.5%; Score 132; DB 3; Length 58;
Best Local Similarity 45.3%; Pred. No. 5,1e-09;
Matches 24; Conservative 9; Mismatches 20; Indels 0

Qy 39 DPCKLDMFGSCYEYHFRFYFNFTSKRCETFFVSGCGNGLNNFKLKIEREVAC 91
| | : : | | | | | : | | | | : | | : | | : |
Db 3 DCFCLFEDTGPCRGDRFYNNQTQCEFFIYGCBCGNMNFETLECKNIC 55

RESULT 12
US-08-676-124-77
; Sequence 77, Application US/08676124
; Patent No. 6010880
; GENERAL INFORMATION:
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert Charles
; TITLE OF INVENTION: INHIBITORS OF HUMAN PLASMIN DERIVED

```

```

RESULT 14
US-09-240-136-77
; Sequence 77, Application US/09240136
; Patent No. 6103499
; GENERAL INFORMATION:
; APPLICANT: DVAX CORP
; APPLICANT: MARKLAND, William
; APPLICANT: LADNER, Robert C
; TITLE OF INVENTION: Inhibitors of Human Plamin Derived
; TITLE OF INVENTION: From the Kunitz Domains
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yankwich & Associates
; STREET: 130 Bishop Allen Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5-inch diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Microsoft Windows 98
; SOFTWARE: Microsoft Word 97 SR-1

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:34 ; Search time 25.25 Seconds
(without alignments)
448.920 Million cell updates/sec

Title: US-09-215-435-167

Perfect score: 1878

Sequence: 1 MYPFIYLAHFHLCGSWSST.....EGPFCHLGHRLWLQHSSTR 351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1673	89.1	501	1	BACE_HUMAN
2	1661	88.4	501	1	BACE_MOUSE
3	1655	88.1	501	1	BACE_RAT
4	925	49.3	518	1	BAE2_HUMAN
5	259	13.8	324	1	PEP1_GADMO
6	253.5	13.5	367	1	PEP1_CHICK
7	251	13.4	388	1	PEP2_MACFU
8	239	12.7	387	1	PEP1_RABIT
9	237.5	12.6	387	1	PEP3_RABIT
10	234.5	12.5	388	1	PEP2_MACFU
11	234	12.5	386	1	PEP2_PIG
12	233	12.4	398	1	CATE_RAT
13	231.5	12.3	388	1	PEP1_HUMAN
14	231.5	12.3	388	1	PEP1_MACMU
15	230.5	12.3	383	1	PEPE_CHICK
16	230	12.2	387	1	PEP2_RABIT
17	230	12.2	396	1	CATE_HUMAN
18	229.5	12.2	388	1	PEP1_MACFU
19	229.5	12.2	391	1	CATE_CAVPO
20	229	12.2	387	1	PEP1_RABIT
21	227	12.1	381	1	CHYM_SHEEP
22	224	11.9	412	1	CATD_HUMAN
23	223.5	11.9	388	1	PAG_HORSE
24	220.5	11.7	388	1	PEPE_RABIT
25	220	11.7	386	1	CATE_RABIT
26	220	11.7	397	1	CATE_MOUSE
27	219	11.7	381	1	CHYM_BOVIN
28	218	11.6	377	1	PEPC_MACFU
29	216.5	11.5	407	1	CATD_RAT
30	212	11.3	410	1	CATD_MOUSE
31	210	11.2	398	1	CATD_CHICK
32	209	11.1	388	1	PEPC_HUMAN
33	202	10.8	394	1	PEPC_CAVPO

34	200	10.6	419	1	CARV_CANAL
35	197.5	10.5	345	1	CATD_PIG
36	196	10.4	376	1	PAG2_BOVIN
37	192.5	10.3	389	1	PAG1_PIG
38	184.5	9.8	402	1	REN1_RAT
39	183.5	9.8	392	1	PEPC_RAT
40	183.5	9.8	420	1	PAG2_PIG
41	182	9.7	380	1	PAG1_BOVIN
42	178.5	9.5	387	1	ASPP_AEDAE
43	178	9.5	396	1	CARP_NEUCR
44	175.5	9.3	394	1	PEPA_ASFAW
45	175	9.3	496	1	ASPR_ORYSA

ALIGNMENTS

RESULT 1
BACE_HUMAN
ID BACE_HUMAN STANDARD; PRT: 501 AA.
AC P56817: Q9UJT5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
DE (MEMAPIN-2);
GN BACE OR BACE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA Treanor J., Rogers G., Citron M.;
RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT the transmembrane aspartic protease BACE.";
RN Science 286:735-741(1999).
RP [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 46-68, AND CHARACTERIZATION
RC TISSUE-BRAIN;
RX MEDLINE=20057171; PubMed=10591214;
RA Sinha S., Anderson J.P., Barbour R., Basi G.S., Caccavello R.,
RA Davis D., Doan M., Doney H.F., Frigon J., Hong J., Jacobson-Croak K.,
RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,
RA Tatsuno G., Tung J., Schenk D., Seubert P., Smeetsaert S.M., Wang S.,
Walker D., Zhao J., McConlogue L., Varghese J.;
RT "Purification and cloning of amyloid precursor protein beta-secretase
RT from human brain.";
RN Nature 402:537-540(1999).
RP [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057170; PubMed=10591213;
RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA Brashler J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-
RT secretase activity.";
RN Nature 402:533-537(1999).
RP [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20030166; PubMed=10561122;
RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
RA Chapman C., Glover I.S., Murphy K.E., Southan C.D., Ryan D.M.,
RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;
RT "Identification of a novel aspartic proteinase (Asp 2) as beta-
RT secretase.";

Mol. Cell. Neurosci. 14:419-427(1999).

[5]

SEQUENCE OF 14-501 FROM N.A., AND CHARACTERIZATION.

MEDLINE=20144060; PubMed=10677483;

Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;

"Human aspartic protease memapsin 2 cleaves the beta-secretase site of beta-amyloid precursor protein.";

Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

-!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP. LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.

-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

-!- TISSUE SPECIFICITY: BRAIN.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF190725; AAF04142.1; -

EMBL; AF201468; AAF18982.1; -

EMBL; AF200343; AAF17079.1; -

EMBL; AF204943; AAF26367.1; -

EMBL; AF200193; AAF13715.1; -

MIN; 604252; -

INTERPRO; IPR001461; -

PFAM; PF00026; asp; 3.

PRINTS; PR00792; PEPsin.

PROSITE; PS00141; ASP_PROTEASE; 1.

Hydrolase; Aspartyl protease; Zymogen; Transmembrane; Signal.

FT SIGNAL 1 21 POTENTIAL.

FT PROPEP 22 45

FT CHAIN 46 501

FT DOMAIN 22 457

FT TRANSMEM 458 478

FT DOMAIN 479 501

FT ACT_SITE 93 93

FT ACT_SITE 289 289

FT CARBOHYD 153 153

FT CARBOHYD 172 172

FT CARBOHYD 223 223

FT CARBOHYD 354 354

FT CARBOHYD 501 AA; 55763 MW; 377CEAC824ACEF05 CRC64;

SEQUENCE 501 AA; 55763 MW; 377CEAC824ACEF05 CRC64;

Query Match 89.1%; Score 1673; DB 1; Length 501;

Best Local Similarity 95.3%; Pred. No. 1.4e-137;

Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;

QY 3 PTI--YLOAHTLCGWSSTYDLKRGVVPYTOGKWEGLGTDLVSPHGNVTVRANI 60

DB 107 PFLHYRYQQL-----SSTYDLKRGVVPYTOGKWEGLGTDLVSPHGNVTVRANI 160

QY 61 AATSDKEFFINGNWEGTLGLAYAEIARPDSPPEFFDSLVKQTHVNLFLSLQLCGAGF 120

DB 161 AATSDKEFFINGNWEGTLGLAYAEIARPDSPPEFFDSLVKQTHVNLFLSLQLCGAGF 220

QY 121 PLNQSEVLASVSGSMIGGIDHSLYTGSLWYTPIRREWYEVIIYVRVEINGODLKMCKE 180

DB 221 PLNQSEVLASVSGSMIGGIDHSLYTGSLWYTPIRREWYEVIIYVRVEINGODLKMCKE 280

QY 181 YNYDKSIVDSGTTNLRPLPKKVFEEAVKSIKAASSTKPDGFWLGEOLVCWQAGTTPWNI 240

DB 281 YNYDKSIVDSGTTNLRPLPKKVFEEAVKSIKAASSTKPDGFWLGEOLVCWQAGTTPWNI 340

QY 241 FPVISLYLMGEVNTQSFRTITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300

DB 341 FPVISLYLMGEVNTQSFRTITLPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 400

QY 301 GFVVVFDRAKRGFAVSACHVHDEFTAAVEGPFCHL 338

DB 401 GFVVVFDRAKRGFAVSACHVHDEFTAAVEGPFVTL 438

RESULT 2

BACE_MOUSE

ID BACE_MOUSE STANDARD; PRT: 501 AA.

AC P56818;

DT 30-MAY-2000 (Rel. 39, Created)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE BETA-SECRETASE PRECURSOR (EC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)

DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2) (MEMAPSIN-2).

DE BACE.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20002972; PubMed=10531052;

RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A., Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y., Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A., Blere A.L., Curran E., Burgess T., Louis J.-C., Collins F., Treanor J., Rogers G., Citron M.;

RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.";

RN Science 286:735-741(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20057170; PubMed=10591213;

RA Van R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M., Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B., Tomasselli A.G., Farodi L.A., Heinrikson R.L., Gurney M.E.;

RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity.";

RL Nature 402:533-537(1999).

CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP. LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- TISSUE SPECIFICITY: BRAIN.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE EUKARYOTIC ASPARTYL PROTEASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF190726; AAF04143.1; -

EMBL; AF200346; AAF17082.1; -

MGD; MGI:1346542; BACE.

INTERPRO; IPR001461; -

PFAM; PF00026; asp; 3.

PRINTS; PR00792; PEPsin.

PROSITE; PS00141; ASP_PROTEASE; 1.

Hydrolase; Aspartyl protease; Zymogen; Transmembrane;

KW Signal. 1 21
 FT PROPEP 22 45
 FT CHAIN 46 501
 FT DOMAIN 22 457
 FT TRANSMEM 458 478
 FT DOMAIN 479 501
 FT ACT_SITE 93 93
 FT ACT_SITE 289 289
 FT CARBOHYD 153 153
 FT CARBOHYD 172 172
 FT CARBOHYD 223 223
 FT CARBOHYD 354 354
 FT CONFLICT 6 6
 FT CONFLICT 81 87
 SQ SEQUENCE 501 AA; 55881 MW; B2FD346419376ACF CRC64;

Query Match 88.4%; Score 1661; DB 1; Length 501;
 Best Local Similarity 94.3%; Pred. No. 1.5e-136;
 Matches 316; Conservative 5; Mismatches 6; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCSGWSSTYRDLRGVVPYPTQKGWEGELGTDLSVPHGPNVTVRANI 60
 DB 107 PFLHRYQRQL-----SSTYRDLRGVVPYPTQKGWEGELGTDLSVPHGPNVTVRANI 160

QY 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLFSLQLCGAGF 120
 DB 161 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLFSLQLCGAGF 220

QY 121 PLNQSEVLASVGSMSIIGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKE 180
 DB 221 PLNQTEALASVGSMSIIGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKE 280

QY 181 YNVDKSIDVSGTNTNLRPKKVEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
 DB 281 YNVDKSIDVSGTNTNLRPKKVEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 340

QY 241 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 300
 DB 341 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 400

QY 301 GFYVVDRAKRKRIGFAVSACHVHDEFRTAAVEGPF 335
 DB 401 GFYVVDRAKRKRIGFAVSACHVHDEFRTAAVEGPF 435

RESULT 3
 BACE_RAT
 ID BACE_RAT STANDARD; PRT; 501 AA.
 AC P56819;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE BETA-SECRETASE PRECURSOR (RC 3.4.23.-) (BETA-SITE APP CLEAVING ENZYME)
 DE (BETA-SITE AMYLOID PRECURSOR PROTEIN CLEAVING ENZYME) (ASPARTYL
 DE PROTEASE 2) (ASP 2) (ASP2) (MEMBRANE-ASSOCIATED ASPARTIC PROTEASE 2)
 DE (MEMAPSN-2).
 GN BACE.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20002972; PubMed-10531052;
 RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
 RA Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
 RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
 RA Biere A.L., Curran E., Burgess T., Louis J.C., Collins F.,
 RA Treanor J., Rogers G., Citron M.;
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
 RT the transmembrane aspartic protease BACE.";
 RL Science 286:735-741(1999).

CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF190727; AAF04144.1; -.
 CC INTERPRO: IPR001461; -.
 CC PFAM: PF00026; asp; 3.
 CC PRINTS: PRO0792; PEPsin.
 CC PROSITE: PS00141; ASP_PROTEASE; 1.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
 KW Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 45 POTENTIAL.
 FT CHAIN 46 501 BETA-SECRETASE.
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 458 478 POTENTIAL.
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 501 AA; 55806 MW; 24B445BC8BE87DE3 CRC64;

Query Match 88.1%; Score 1655; DB 1; Length 501;
 Best Local Similarity 94.0%; Pred. No. 5e-136;
 Matches 315; Conservative 5; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCSGWSSTYRDLRGVVPYPTQKGWEGELGTDLSVPHGPNVTVRANI 60
 DB 107 PFLHRYQRQL-----SSTYRDLRGVVPYPTQKGWEGELGTDLSVPHGPNVTVRANI 160

QY 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLFSLQLCGAGF 120
 DB 161 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFDSLKVQTHVPNLFSLQLCGAGF 220

QY 121 PLNQSEVLASVGSMSIIGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKE 180
 DB 221 PLNQTEALASVGSMSIIGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKE 280

QY 181 YNVDKSIDVSGTNTNLRPKKVEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
 DB 281 YNVDKSIDVSGTNTNLRPKKVEAAVKSIAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 340

QY 241 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 300
 DB 341 FVVISLYLMGEVNTQSFRTILPQOYLPRVEDVATSDDDCYKFAISQSSTGTVMGAVIME 400

QY 301 GFYVVDRAKRKRIGFAVSACHVHDEFRTAAVEGPF 335
 DB 401 GFYVVDRAKRKRIGFAVSACHVHDEFRTAAVEGPF 435

RESULT 4
 BAE2_HUMAN
 ID BAE2_HUMAN STANDARD; PRT; 518 AA.

[illegible]

DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PEP4_CHICK (EC 3.4.23.-)
 OS Gadus morhua (Atlantic cod).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidae; Gadidae;
 OC Gadus.
 [1]
 RN SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
 RP TISSUE-STOMACH;
 RC Karlsten S., Hough E., Olsen R.L.;
 RA "Structure and proposed amino-acid sequence of a pepsin from Atlantic
 RT cod (Gadus morhua).";
 RL Acta Crystallogr. D 54:32-46(1998).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 DR PDB; 1AM5; 24-DEC-97.
 DR INTERPRO: IPR001461; -
 DR INTERPRO: IPR001969; -
 DR PFAM; PF00026; asp: 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Gastric juice;
 KW 3D-structure.
 FT ACT_SITE 32 32 BY SIMILARITY.
 FT ACT_SITE 214 214 BY SIMILARITY.
 FT DISULFID 45 50 BY SIMILARITY.
 FT DISULFID 206 209 BY SIMILARITY.
 FT DISULFID 247 280 BY SIMILARITY.
 SQ SEQUENCE 324 AA; 34014 MW; EE3A6097B6941DD7 CRC64;

Query Match 13.8%; Score 259; DB 1; Length 324;
 Best Local Similarity 28.7%; Pred. No. 3.3e-15;
 Matches 90; Conservative 48; Mismatches 114; Indels 62; Gaps 14;
 QY 18 SSTYRDLRKGVVYPTQCKWEGELGTLVSIPIHG--PNVTVRANIAAITESDKFFINGSN 75
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 61 SSIYVETGKTVLTYGTGNGRILGQDTSVGGSDPNQELG---ESQTEPGPQA-AAP 116
 QY 76 WEGILGAYAEIARPDSPPEFFDSLVKQTHV-PNIFSLQLCGAGPPLNOSSEVLASVGS 134
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 117 FDGILGLAYPSIAAA--GAVPVEDNMGSSQLVEKDLFSYLSGGG--ANGSEVM----- 166
 QY 135 MITGGIDHSYLTSLWTPTRREWEYVEIIVRVEINGODLKM-DCKEYNDKSIDSGTT 193
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 167 --LGGVDNSHYTGSIIHWIPVTAEKYQWALDGTIVNGQTAAECGC-----QAIVDGTGS 218
 QY 194 NLRPLPKVFEAAVKSIAASSTKEKFPDGFGLQVLCWQAGTTPWNIFPVISLYLMGEVT 253
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 219 KIVAPVSALANINKIGASEN-----QGEMWGN--CASVQSLPDITF-----T 259
 QY 254 NQSFRTILPQOYLRPVEDVATSDCCYKFAISOSSTGT-----VMGAVIMEGFYVV 305
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 260 INGKQPLPSAYIEGQAFCTSD-----GLGSGVPSNTSELWIFGDVFLRNYITI 310
 QY 306 FDRARKRIGFAVSA 319
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 311 YDRTNKVKGFAPAA 324

RESULT 6
 PEP4_CHICK
 ID PEP4_CHICK STANDARD; PRT; 367 AA.
 AC P00793;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEPsin A PRECURSOR (EC 3.4.23.1).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=84004412; PubMed=6617663;
 RA Baudys M., Kostka V.;
 RT "Covalent structure of chicken pepsinogen.";
 RL Eur. J. Biochem. 136:89-99(1983).
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 CC ALSO CLEAVED TO SOME EXTENT.
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
 CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
 DR PIR; A00984; PECH.
 DR HSSP; P00794; 3CMS.
 DR MEROPS; A01.001; -
 DR INTERPRO: IPR001461; -
 DR INTERPRO: IPR001969; -
 DR PFAM; PF00026; asp: 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Stomach; Zymogen;
 KW Glycoprotein; Gastric juice.
 FT PROPEP 1 42 ACTIVATION PEPTIDE.
 FT CHAIN 43 367 PEPsin A.
 FT ACT_SITE 77 77
 FT ACT_SITE 260 260
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).
 FT DISULFID 90 95
 FT DISULFID 251 255
 FT DISULFID 290 323
 SQ SEQUENCE 367 AA; 40431 MW; 0C547E7FD8F5B341 CRC64;

Query Match 13.5%; Score 253.5; DB 1; Length 367;
 Best Local Similarity 23.6%; Pred. No. 1.2e-14;
 Matches 80; Conservative 62; Mismatches 114; Indels 83; Gaps 13;
 QY 2 VPFTYLOAHETLCSGW-----SSTYRDLRKGVVYPTQCKWEGELGTLVSIPIHGPNV 54
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 85 VPSIYCKS--SACSNHRRFDPKSSSTVSTNETVYIAYTGSMGILGYDTPVAV---SSI 139
 QY 55 TVRANIAAITESDK-PFINGSNWEGILGAYAEIARPDSPPEFFDSLVKQTHV-PNIFLS 112
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 140 DVQNIQFGLSETEPGSFYFCNFDGILGLAAPSIS--SSGATPVFDNMMSHQLVAQDLFS 197
 QY 113 LQLCGAGPPLNQSEVLASVGSMTIIGDHSYLTGSLWYTPIRREWEYVEIIVRVEINGQ 172
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 198 VYLSKDG-----ETGSFVLFGGIDPNYTTKGIYVWPLSAETIYQITMDRVTVGNK 247
 QY 173 DLK--MDCKEYNDKSIDSGTTNLRPKVFEAAVKSIAASSTE-----KFPDGF 223
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 248 YVACFFTC-----QAIVDGTSLVMPQGAYNRIIKDLGVSSDGEISCDISKLDP--- 298
 QY 224 LGEOLVCWQAGTTPWNIFPVISLYLMGEVNTQSPRITLPPQYLRPVEDVATSDCCYKF 283
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 299 -----VTFHINGHA-----FTLPASAYVLNEDGSCML 325
 QY 284 AISOSSTGT-----VMGAVIMEGFYVFDRAKRIGFA 316
 Db : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 326 GFENMGTPTELGEQWILGDVFIREFYVIFDRANKVGLS 364

RESULT 7
 PEP4_MACFU
 ID PEP4_MACFU STANDARD; PRT; 388 AA.
 AC P27678;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PEPsin A-4 PRECURSOR (EC 3.4.23.1) (PEPSIN I/II).
 GN PGA.
 OS Macaca fuscata fuscata (Japanese macaque).


```

Db 177 PFDGILGLAYPSISASDAT--PVFDNNWNEGLVSEDLFSVYLSNG-----EKGS 224
QY 134 SMIIIGGDHSLYTGSLWYTPIRREWYEVILVRVEINGDQWK--DCKEYNDKSIYDSG 191
Db 225 VMVFGGIDSSYYTGSLWNVPSVHEGYWQITMDSITTINGETIACADSC-----QAVYDTG 278
QY 192 TTNLRPKKVFEEAAVKSIAKASTKPPDGFWMGEQLV-CWQAGTTPWNFFPVSISLYLMG 250
Db 279 TSLLAGPTSIAIKSIQSYGASKNL-----LGENIISCAIDSLPDIVF----- 321
QY 251 EVNQSPRITILPOQYLRPVED--VATSQDDC-----YKPAISQSSTGT--VMGAVIMEGFYV 304
Db 322 -----TINNVPYLPASAVILKDDCLSGFGDMNLDTSYGELWILGDVFIROYFT 372
QY 305 VEDRARKRIGFAVSA 319
Db 373 VEDRANNQVGLAAAA 387

RESULT 9
PEP3_RABIT
ID PEP3_RABIT STANDARD; PRT; 387 AA.
AC P27822;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN III PRECURSOR (EC 3.4.23.1) (PEPSIN A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOSINS AT LATE POSTNATAL STAGE.
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59237; AAA85370.1; -
CC PIR; E38302; E38302.
CC HSSP; P00791; 1PSA.
CC MEROPS; A01.001; -.
CC INTERPRO; IPR001461; -.
CC INTERPRO; IPR001969; -.
CC PFAM; PF00026; asp; 1.
CC PRINTS; P00792; PEPSIN.
CC PROSITE; PS00141; ASP_PROTEASE; 2.
CC K0 Hydrolyase; Aspartyl protease; Digestion; stomach; gastric juice;
KW Zymogen; Signal; Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEPSIN III.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).

```

```

FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 41969 MW; 15A59AC81F36F9EF CRC64;

Query Match 12.6%; Score 237.5; DB 1; Length 387;
Best Local Similarity 26.0%; Pred. No. 3.1e-13; Indels 65; Gaps 15;
Matches 87; Conservative 62; Mismatches 121;

QY 2 VPFIYLAHFTLCSGW-----SSTYRDLRKGVYVPYTGQKWEGLGTDLVSIHPGPNV 54
Db 101 VPSYVCSS--AACSVHQNFPEDSSTFQATSESLITYGTGSMTGFLGYDKVKV--GNI 155
QY 55 TVRANIAAITYESDK-FFINGSNWEGILGLAYAEIARDDSPDPFDSLVRKQTHV-PNLFS 112
Db 156 EDTNQIFGLSESPGFLYAPFDGILGLAYPSISSSDAT--PVFDNNWNEGLVSEDLFS 213
QY 113 LQLCGAGFPLNQSEVLASVGSMILGIDHSLYTGSLWYTPIRREWYEVILVRVEINGQ 172
Db 214 VYLSDD-----ESGVVMFGGIDSSYYTGSLNWPVSYEGYWIITLDSITMDGE 263
QY 173 DLKM--DCKEYNDKSIYDVGTTNLRPKKVFEEAAVKSIAKASTKPPDGFWMGEQLVC 230
Db 264 TIACADSC-----QAIVDTGTSLLAGP---TSAISNIQSYIGASENSD---GEMIVS 309
QY 231 WQAGTTPWNFFPVSILYMGVENVNQSFRITILPOQYLRPVEDVATSQDDC-----YKPAIS 286
Db 310 CS-----SMYSLPNIV-----FTINGVQYVPASAYILEEDDACISGFEGMNL 352
QY 287 QSSPTCT--VMGAVIMEGFYVVEDRARKRIGFAVSA 319
Db 353 DTYTGELWILGDVFIROYFTVEDRANNQVGLAAAA 387

RESULT 10
PEP2_MACFU
ID PEP2_MACFU STANDARD; PRT; 388 AA.
AC P27677;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPSIN A-2/A-3 PRECURSOR (EC 3.4.23.1) (PEPSIN III-2/III-1).
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 16-70.
RC TISSUE-GASTRIC MUCOSA;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
CC -I- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE PHE-, LEU-.
CC -I- DEVELOPMENTAL STAGE: PEP A-2 IS PREDOMINANT AT THE 4-MONTH STAGE.
CC PEP A-3 IS PREDOMINANT AT FETAL STAGES.
CC -I- PTM: PEP A-2 IS PHOSPHORYLATED, BUT NOT PEP A-3.
CC -I- PTM: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING PEPSIN AT PH
CC 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA ACTIVATION
CC SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC CLEAVAGE
CC VIA AN INTERMEDIATE FORM(S).
CC -I- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE
CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.
CC -I- CAUTION: IT IS NOT KNOWN IF THIS IS PEP A-2 OR PEP A-3.

```


Db 355 SGELWILGDVFIQRYTVFDRANKVGLA 383

RESULT 12

CATE_RAT

ID CATE_RAT STANDARD; PRT; 398 AA.

AC P16228; Q63701;

DT 01-APR-1990 (Rel. 14, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE CATHEPSIN E PRECURSOR (EC 3.4.23.34).

GN CRSE.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Rattus.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SPLEEN;

RC MEDLINE=96004785; PubMed=7574663;

RA Okamoto K., Yu H., Misumi Y., Ikehara Y., Yamamoto K.;

RT "Isolation and sequencing of two cDNA clones encoding rat spleen

RT cathepsin E and analysis of the activation of purified procathepsin

RT E.";

RL Arch. Biochem. Biophys. 322:103-111(1995).

RN [2]

RP SEQUENCE OF 59-110.

RX MEDLINE=90147750; PubMed=2105725;

RA Yonezawa S., Takahashi T., Ichinose M., Miki K., Tanaka J., Gasa S.;

RT "Structural studies of rat cathepsin E: amino-terminal structure and

RT carbohydrate units of mature enzyme.";

RL Biochem. Biophys. Res. Commun. 166:1032-1038(1990).

CC -!- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN

CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.

CC -!- CATALYTIC ACTIVITY: SIMILAR TO CATHEPSIN D, BUT SLIGHTLY BROADER

CC SPECIFICITY.

CC -!- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL; D45187; BAA08128.1; -;

DR EMBL; D38104; BAA07285.1; -;

DR PIR; A34657; A34657.

DR PIR; B34643; B34643.

DR PIR; C34643; C34643.

DR HSP; P00794; 3CMS.

DR MEROPS; A01.010; -;

DR INTERPRO; IPR001461; -;

DR INTERPRO; IPR001969; -;

DR PFAM; PF00026; asp. 1.

DR PRINTS; P000792; PEPsin

DR PROSITE; PS00141; ASP_PROTEASE; 2.

DR Hydrolyase: Aspartyl protease; Glycoprotein; Zymogen; Signal;

KW Alternative splicing.

FT SIGNAL 1 19

FT PROPEP 20 58

FT CHAIN 59 398

FT MOD_RES 20 20

FT ACT_SITE 98 98

FT ACT_SITE 283 283

FT DISULFID 62 62

FT DISULFID 111 116

FT DISULFID 274 278

FT CARBOHYD 92 92

FT VARSPLIC 312 344

FT VARIANT 114 114 P -> S.

FT CONFLICT 79 E -> N (IN REF. 2).

FT CONFLICT 84 TV -> SR (IN REF. 2).

SQ SEQUENCE 398 AA; 43021 MW; 25F123E67C46EB5F CRC64;

Query Match 12.4%; Score 233; DB 1; Length 398;

Best Local Similarity 25.4%; Pred. No. 7.8e-13;

Matches 84; Conservative 59; Mismatches 130; Indels 58; Gaps 14;

QY 2 VPFYI-----LQAHFTLCGSGSTYRDLRKGVVYPTQCKWEGELGTDLVSPHGPNTV 56

Db 106 VPSVYCTSPACKAHPVFPSPQSTYMEVGNHFSIQYGTGSLTGIIIGADQVSV-EGLTVEG 164

QY 57 RANIAITSDREFFINGSNWEGTLGLAYAEIARPDSPPEFDSLVKQTHVP-NLFSQL 115

Db 165 QQFGESVKEPGQTFVN-AEFDGLGLGYPSLA--VGVTPVPFDNMAQNVALPHEVYL 221

QY 116 CGAGFPLNQSEVLASVGGSMIIIGDHSLYTGLSWYTPIRREMYEVIIVRVEINGQDLK 175

Db 222 -----SSDPQGGSGSELTTGGYDPSHFSGSLNIPVTQGYMQLDGIQVG--DTV 271

QY 176 MCKEYNYDKSIVDSGTTNLRPLPKVFEAAVKSIAKASSTEFPPDGFGLGQLVCWQAGT 235

Db 272 MFCSE--GCQAIVDGTSLTGPVK-----KIKOLQEA-----IGA 305

QY 236 TPWNIPFISLYLMGEVTNQSFR-----TILPQQYLRPVEDVATSDDC---YKFAISQ 287

Db 306 TPMDGEYVDCATLNMMPNVTFLNGSVYTLSTAVILP-DLVDGMQFGSGFGQLDIO 363

QY 288 SSTGT--VMGAVIMEGVYVDFRARRKIGFA 316

Db 364 PPAGPLWILGDVFIQRYTVFDRANKVGLA 394

RESULT 13

PEPA_HUMAN

ID PEPA_HUMAN STANDARD; PRT; 388 AA.

AC P00790;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE PEPsin A PRECURSOR (EC 3.4.23.1).

GN PGA3 AND PGA4 AND PGAS.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=83161158; PubMed=6300126;

RA Sogawa K., Fujii-Kuriyama Y., Mizukami Y., Ichihara Y., Takahashi K.;

RT "Primary structure of human pepsinogen gene.";

RL J. Biol. Chem. 258:5306-5311(1983).

RN [2]

RP SEQUENCE FROM N.A. (ISOZYME 5).

RC TISSUE=PLACENTA;

RX MEDLINE=89233110; PubMed=2714789;

RA Evers M.P.J., Zelle B., Bebelman J.P., van Beusechem V., Kraakman L.,

RA Hoffer M.J.V., Pronk J.C., Mager W.H., Planta R.J., Eriksson A.W.,

RA Frants R.R.;

RT "Nucleotide sequence comparison of five human pepsinogen A (PGA)

RT genes: evolution of the PGA multigene family.";

RL Genomics 4:232-239(1989).

RN [3]

RP PARTIAL SEQUENCE OF 1-28.

RX MEDLINE=86059312; PubMed=2415509;

RA Ichihara Y., Sogawa K., Takahashi K.;

RT "Isolation of human, swine, and rat prepepsinogens and calf

RT preprochymosin, and determination of the primary structures of their

RL NH2-terminal signal sequences.";

RN J. Biochem. 98:483-492(1985).

RN [4]

RP SEQUENCE OF 16-100 (ISOZYME 2; 3; 3A; 4 AND 5).


```
Db 219 -REFMGSWVFGGIDSEYFTGSINWIPVSYQGYWQISMDSIIVNKQEIACSSGC----- 271
QY 185 KSI VDSGTTNLR LPKKYFEAAVKSIKAASTERPPDGFWLGEQLVCHQAGTTPWNI EPVI 244
Db 272 QAIIDTGTSLVAGPASDINDIQSAVGANQNT-----YGEYSV-----NCSHIL 314
QY 245 SL-----YIMGEVTNQSFRITILPQQYL RPVEDVA-----TSQDDCYKFAISOSSTGTVMGA 296
Db 315 AMPDVVFVIGGI-----QY--PVPALAYTEQNGQGTMCSSSFQNSSADLWILGD 360
QY 297 VIMEGFYVFDRAKRIGFA 316
Db 361 VFIRVYYSIEDRANNRVGLA 380
```

Search completed: March 24, 2001, 13:24:20
Job time: 166 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:29 ; Search time 47.63 seconds
(without alignments)
251.985 Million cell updates/sec

Title: US-09-215-435-167

Perfect score: 1878

Sequence: 1 MWPFIYLAHFTLCGSWSST.....EGPCHLGHRLWLQHSR 351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_36:*

- 1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	100.0	351	20 Y35918	Extended human sec
2	1673	89.1	453	21 Y88438	Modified human asp
3	1673	89.1	459	21 Y88439	Modified human asp
4	1668	88.8	425	21 Y88437	Human Asp2 amino a
5	1668	88.8	433	21 Y88433	Human-pro-Asp-2(a)
6	1668	88.8	446	21 Y88431	T7-caspase-human-p
7	1668	88.8	459	21 Y88432	T7-caspase-human-p
8	1668	88.8	501	21 Y88425	Human aspartyl pro
9	1667	88.8	501	19 W59807	Amino acid sequenc
10	1661	88.4	501	21 Y88427	Murine aspartyl pr
11	1656	88.2	790	19 W59808	Partial amino acid
12	1525.5	81.2	476	21 Y88426	Human aspartyl pro

13	925	49.3	518	19 W61362	Aspartic proteinas
14	925	49.3	518	20 Y41714	Human PRO852 prote
15	925	49.3	518	20 Y22239	Human CSP56, aspar
16	925	49.3	518	20 Y13799	Human aspartyl pro
17	925	49.3	518	21 Y88424	Human aspartyl pro
18	292	15.5	53	20 Y11427	Human 5' EST secre
19	228	12.1	381	13 R20730	Prochymosin (prore
20	224	11.9	412	16 R74207	Human death associ
21	224	11.9	412	19 W71369	Death associated p
22	224	11.9	412	20 Y06478	Human tumour-associ
23	223	11.9	365	4 P30603	Sequence encoded b
24	223	11.9	365	11 R05080	Sequence of calf p
25	223	11.9	375	5 P40078	Sequence encoded b
26	223	11.9	380	3 P20038	Pre-prorennin-A pr
27	223	11.9	381	5 P40559	Sequence of a poly
28	222	11.8	381	5 P40218	Sequence of rennin
29	219	11.7	365	10 P94144	Prochymosin. P94
30	219	11.7	458	10 P94376	BamHI/Sall insert
31	219	11.7	545	20 Y33830	Oleosin-spacer-Met
32	218	11.6	381	4 P30086	Sequence encoded b
33	216	11.5	388	20 Y32058	Cat pregnancy asso
34	215	11.4	381	4 P30446	Sequence encoded b
35	214	11.4	379	4 P30013	Sequence encoded b
36	203	10.8	375	20 Y32055	Bovine pregnancy a
37	199	10.6	391	20 Y32056	Bovine pregnancy a
38	197.5	10.5	376	20 Y32054	Bovine pregnancy a
39	197	10.5	450	10 P94370	Sequence encoded b
40	196.5	10.5	392	20 Y32057	Bovine pregnancy a
41	196	10.4	376	20 Y32035	Bovine pregnancy a
42	193	10.3	380	20 Y32041	Bovine pregnancy a
43	193	10.3	387	20 Y32052	Bovine pregnancy a
44	185.5	9.9	381	20 Y32046	Bovine pregnancy a
45	185	9.9	450	21 Y57041	Plasmodium vivax p

ALIGNMENTS

RESULT	1
Y35918	
ID	Y35918 standard; Protein; 351 AA.
XX	
AC	Y35918;
XX	
XX	
DT	13-SEP-1999 (first entry)
DE	Extended human secreted protein sequence, SEQ ID NO. 167.
XX	
KW	Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW	cellular differentiation; immune system regulator; anti-inflammatory;
KW	haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW	reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
XX	genetic disease.
OS	Homo sapiens.
XX	
PN	W09931236-A2.
XX	
PD	24-JUN-1999.
XX	
XX	
PF	17-DEC-1998; 98WO-IB02122.
XX	
PR	10-AUG-1998; 98US-0096116.
PR	17-DEC-1997; 97US-0069957.
PR	09-FEB-1998; 98US-0074121.
PR	13-APR-1998; 98US-0081563.
XX	
PA	(GEST) GENSET.
XX	
PI	Bougueleret L, Duclert A, Dumas Milne Edwards J;
DR	WPI: 1999-385906/32.
DR	N-PSDB; X97602.

```
XX PT New isolated human secreted proteins
XX PS Claim 9; Page 198-199; 516pp; English.
XX CC This sequence is encoded by an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases.
XX Sequence 351 AA;
SQ
Query Match 100.0%; Score 1878; DB 20; Length 351;
Best Local Similarity 100.0%; Pred. No. 9.9e-197;
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYPFIYQLAHFTLCGWSSTYRDLRKGVVYPTQCKWEGELGTDLSVPHGPNVTVRANI 60
DB 1 mvpfiylqahftlcgwsstydrlrkgvvpytgqkwegeigtdivsphgpnvtvrani 60
QY 61 AATFSDKFFINGNWEGILGLAYAEIARPDSPFPFDSLVKQTHVPLNLSLQLCGAGF 120
DB 61 aaatesdkffingsnwegilglayaeiarpdspfpfddslvkqthvplnlsqldcge 120
QY 121 PLNQSEVLASVGGSMIIGIDHSLYTGSLWYTPPIREWYVEIIVRVEINGDLKMDCKE 180
DB 121 plnqsevlasvgsmligdhsltygslwytppirrewyveilvrveingdldkmdcke 180
QY 181 YNYDKSIVDSGTTNLRPKKVFEEAVKSTKAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 240
DB 181 ynydksivdsdgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwagttpwni 240
QY 241 FPVISLYLMEVTNQSFRTITLPQOYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIME 300
DB 241 fpvislylmevtnqsfritilpqoylrpvedvatssqddcykfaissqstgtvmgavime 300
QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPCHLGHRLWLOHSTDR 351
DB 301 gfyvvfdrarkrigfavsachvhdefrtaavegpchlghgrlwlqhstdr 351
RESULT 2
Y88438
ID Y88438 standard; Protein; 453 AA.
XX AC
XX Y88438;
XX DT
XX 03-AUG-2000 (first entry)
XX DE Modified human aspartyl protease 2 (Asp2) amino acid sequence.
XX KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
XX KW Alzheimer's disease; beta secretase site.
XX OS Homo sapiens.
XX PN WO200017369-A2.
XX PD
XX 30-MAR-2000.
XX PF 23-SEP-1999; 99WO-0520881.
XX XX
XX 24-SEP-1998; 98US-0101594.
XX PR
```

```
XX PA (PHAA ) PHARMACIA & UPJOHN CO.
XX PI Gurney ME, Bienkowski MJ, Helnrirkson RL, Parodi LA, Yan R;
XX DR N-PSDB; A15688.
XX WPI: 2000-303209/26.
XX PT New enzyme designated human aspartase useful in research into
XX Alzheimer's disease is capable of cleaving amyloid protein precursor at
XX the beta secretase site to produce amyloid beta peptide
XX Example 10; Page 169-172; 183pp; English.
XX This sequence represents a modified human aspartyl protease 2 (Asp2)
XX amino acid sequence. Asp2 encoded by this sequence has the C-terminal
XX transmembrane domain deleted. The invention relates to a protease
XX (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
XX precursor protein (APP). The protease contains a sequence encoding the
XX amino acid sequence DTG and a sequence encoding DSG or DTG separated by
XX 100-300 amino acids. When mutated the APP gene causes an autosomal
XX dominant form of Alzheimer's disease. APP localises to the cell surface
XX membrane and have a single C-terminal transmembrane domain. Proteolytic
XX processing of APP produces the amyloid beta protein, which is possibly
XX very important in Alzheimer's disease. The invention includes a
XX nucleotide sequence encoding the protease, a vector containing the
XX screening for inhibitors of beta secretase activity are also given in the
XX invention. The human aspartase protein and nucleotide sequences and the
XX methods for identifying inhibitors of the protease, are useful in the
XX treatment of and research in to Alzheimer's disease.
XX Sequence 453 AA;
SQ
Query Match 89.18; Score 1673; DB 21; Length 453;
Best Local Similarity 95.3%; Pred. No. 3.7e-174;
Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;
QY 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQCKWEGELGTDLSVPHGPNVTVRANI 60
DB 107 pfihryygrql-----sstydrirkgvvpytgqkwegeigtdivsphgpnvtvrani 160
QY 61 AATFSDKFFINGNWEGILGLAYAEIARPDSPFPFDSLVKQTHVPLNLSLQLCGAGF 120
DB 161 aaatesdkffingsnwegilglayaeiarpdspfpfddslvkqthvplnlsqldcge 220
QY 121 PLNQSEVLASVGGSMIIGIDHSLYTGSLWYTPPIREWYVEIIVRVEINGDLKMDCKE 180
DB 221 plnqsevlasvgsmligdhsltygslwytppirrewyveilvrveingdldkmdcke 280
QY 181 YNYDKSIVDSGTTNLRPKKVFEEAVKSTKAASSTKPEPDGFWLGEQLVCWQAGTTPWNI 240
DB 281 ynydksivdsdgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwagttpwni 340
QY 241 FPVISLYLMEVTNQSFRTITLPQOYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIME 300
DB 341 fpvislylmevtnqsfritilpqoylrpvedvatssqddcykfaissqstgtvmgavime 400
QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPCHL 338
DB 401 gfyvvfdrarkrigfavsachvhdefrtaavegpchl 438
RESULT 3
Y88439
ID Y88439 standard; Protein; 459 AA.
XX AC
XX Y88439;
XX DT
XX 03-AUG-2000 (first entry)
XX XX
XX Modified human aspartyl protease 2 (Asp2) amino acid sequence.
XX DE
```

XX Asparyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
OS Homo sapiens.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US20881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
DR N-PSDB; A15689.
XX
PT New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 10; Page 173-176; 183pp; English.
PS
CC This sequence represents a modified human aspartyl protease 2 (Asp2)
CC amino acid sequence. Asp2 encoded by this sequence has the C-terminal
CC transmembrane domain deleted. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
SQ Sequence 459 AA;

Query Match 89.1%; Score 1673; DB 21; Length 459;
Best Local Similarity 95.3%; Pred. No. 3.8e-174;
Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQGWEGELGTLVSIHPGPNVTVRANI 60
DB 107 pfhryyqrql-----sstyrdlrgvvyptqgkwegelgtlvsiphgpnvtvrani 160
QY 61 AAITESDKFFINGSNWEGTLGLAYAEIARPDSPFPFDSLKVQTHVNLFSLQLCGAGF 120
DB 161 aaitesdkffingsnwegtlglayaeiarppdslepfdslvkqthvnlfsqlcgagf 220

QY 121 PLNQSEVLASVSGSMIIIGIDHSLTGSLWYTPIRREWYEVIIVRVEINGODLKMDCKE 180
DB 221 plnqsevlasvsgsmiigidhslytgslywtpirreyeviivrveingodlkmcdke 280

QY 181 YNYDKSIVDSGTTNLRPKKVFEEAAVKSTKAASSTKFPDGFWMGEQLVCWQAGTTPWNI 240
DB 281 ynydksivdsgettlnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagttwni 340

QY 241 FPIVSLYLMGEVTNOSFRITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
DB 341 fpvislylmgevtngsfrilitlpqoylprvedvatsddcykfaissqstgtvmgavime 400

QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVEGPTFCHL 338

DB 401 gfyyvfdarkrigrfavsachvhdefrtaavegpfvtl 438
RESULT 4
Y88437
ID Y88437 standard; Protein; 425 AA.
XX
AC Y88437;
XX
DT 03-AUG-2000 (first entry)
XX
DE Human Asp2 amino acid sequence containing proteolytic cleavage site.
XX
KW Asparyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site.
XX
OS Homo sapiens.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US20881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
PI Gurney ME, Bienkowski MJ, Heinrichson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
DR N-PSDB; A15677.
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 9; Page 166-168; 183pp; English.
PS
XX This sequence represents a modified version of the human aspartase 2
CC (Asp2) nucleotide sequence. The sequence is used in the bacterial
CC expression of human Asp2L. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
SQ Sequence 425 AA;

Query Match 88.8%; Score 1668; DB 21; Length 425;
Best Local Similarity 95.0%; Pred. No. 1.2e-173;
Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;

QY 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQGWEGELGTLVSIHPGPNVTVRANI 60
DB 79 pfhryyqrql-----sstyrdlrgvvyptqgkwegelgtlvsiphgpnvtvrani 132

QY 61 AAITESDKFFINGSNWEGTLGLAYAEIARPDSPFPFDSLKVQTHVNLFSLQLCGAGF 120
DB 133 aaitesdkffingsnwegtlglayaeiarppdslepfdslvkqthvnlfsqlcgagf 192

QY 121 PLNOSVLASVSGSMIIIGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKE 180
DB 193 plngsevlavsgsmiigldhsltygslwypirrewyeviiivrvveingdldmcke 252
QY 181 YNYDKSIIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNI 240
DB 253 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagtpwni 312
QY 241 FVISLYLMGEVNTQSFRTITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
DB 313 fpvislylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavime 372
QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVGPFCHL 338
DB 373 gfyvvdrrarkrigfavsachvhdefrtaavegpfvtl 410

RESULT 5
Y88433
ID Y88433 standard; Protein; 433 AA.
XX
AC Y88433;
XX
DT 03-AUG-2000 (first entry)
XX
DE Human-pro-Asp-2(a)-deltaTM amino acid sequence.
XX
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site; human-pro-Asp-2(a)-deltaTM.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Bacteriophage t7.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US20881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
DR N-PSDB; A15670.
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 9; Fig 8; 183pp; English.
XX
XX This sequence represents a modified version of the human aspartase 2
CC (Asp2) amino acid sequence. The sequence is used in the bacterial
CC expression of human Asp2L. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
XX Sequence 433 AA;

Query Match 88.8%; Score 1668; DB:21; Length 433;
Best Local Similarity 95.0%; Pred. No. 1.2e-173;
Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
QY 3 PFI--YLQAHFTLCSGMSSTVRLRKGVYVPYTOGKWEGLGTLVSIPIHGPNVTVRANI 60
DB 87 pfihryyqrql-----sstyrdirksyvyptqgkwegeigtldvsiipghpnvtrani 140
QY 61 AAITESDKFFINGNWGILGLAYAEIARPDSPDPFDSLVKQTHVPLNLSLQCGAGF 120
DB 141 aaaitesdkffingnwegilglayaeiarpdspdpfdsivkqthvplnlslnlcagaf 200
QY 121 PLNOSVLASVSGSMIIIGIDHSLYTGSLWYTPIRREWYEVIIIVRVEINGDLMCKE 180
DB 201 plngsevlavsgsmiigldhsltygslwypirrewyeviiivrvveingdldmcke 260
QY 181 YNYDKSIIVDSGTTNLRPKKVFEEAAVKSIAASSTKFPDGFWLGEQLVCWQAGTTPWNI 240
DB 261 ynydksivdsgttnlrpkkvfeaaavksikaasstekfpdgfwlgeqlvcwqagtpwni 320
QY 241 FVISLYLMGEVNTQSFRTITILPQOYLPRVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
DB 321 fpvislylmgevtngsfritilpqgylrpvedvatsqdcykfaissgstgtvmgavime 380
QY 301 GFYVVDRAKRIGFAVSACHVHDEFRTAAVGPFCHL 338
DB 381 gfyvvdrrarkrigfavsachvhdefrtaavegpfvtl 418

RESULT 6
Y88431
ID Y88431 standard; Protein; 446 AA.
XX
AC Y88431;
XX
DT 03-AUG-2000 (first entry)
XX
DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.
XX
KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
KW Alzheimer's disease; beta secretase site;
KW T7-caspase-human-pro-Asp-2(a)-deltaTM.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Bacteriophage t7.
XX
PN WO200017369-A2.
XX
PD 30-MAR-2000.
XX
PF 23-SEP-1999; 99WO-US20881.
XX
PR 24-SEP-1998; 98US-0101594.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
XX
XX WPI; 2000-303209/26.
DR N-PSDB; A15668.
XX
XX New enzyme designated human aspartase useful in research into
PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
PT the beta secretase site to produce amyloid beta peptide -
XX
XX Example 9; Fig 6; 183pp; English.
XX
XX This sequence represents a modified version of the human aspartase 2
CC (Asp2) amino acid sequence. The sequence is used in the bacterial
CC expression of human Asp2L. The invention relates to a protease
CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
CC precursor protein (APP). The protease contains a sequence encoding the
CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
CC 100-300 amino acids. When mutated the APP gene causes an autosomal
CC dominant form of Alzheimer's disease. APP localises to the cell surface
CC membrane and have a single C-terminal transmembrane domain. Proteolytic
CC processing of APP produces the amyloid beta protein, which is possibly
CC very important in Alzheimer's disease. The invention includes a
CC nucleotide sequence encoding the protease, a vector containing the
CC nucleotide sequence, and a cell line comprising the vector. Methods for
CC screening for inhibitors of beta secretase activity are also given in the
CC invention. The human aspartase protein and nucleotide sequences and the
CC methods for identifying inhibitors of the protease, are useful in the
CC treatment of and research in to Alzheimer's disease.
XX
XX Sequence 433 AA;

CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 SQ Sequence 446 AA;

Query Match 88.8%; Score 1668; DB 21; Length 446;
 Best Local Similarity 95.0%; Pred. No. 1.3e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
 Qy 3 PFI--YLOAHFTLCGWSSTYRDLRKGVYVPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
 Db 100 pflhryyqrql-----sstyrdlrkgyvpytgkgwgelgtdlvsiphgpnvtvrani 153
 Qy 61 AAITESDRFFINGSNWEGILGLAYAEIARPDSPFPFDSLKQTHVPLNLSLQLCGAGF 120
 Db 154 aaatesdkffingsnwegilglayaeiarpdpslepfdslykqthvplnlslhlcgagf 213
 Qy 121 PLNQSEVLASVGGSMIIGGIDHSlyTGSlyWTPPIRREWYVEIIVRVEINGODLKMDCKE 180
 Db 214 plnqsevlasvggsmilggidhslytgslywtpirrewyveilvrveingodlkmdccke 273
 Qy 181 YNDKSIIVDSGTTNLRPKKVFEEAAVKSIAASSTKEKFPDGFGLGEQLVLCWQAGTTPWNI 240
 Db 274 ynydksivdsdgttnlrpkpkvfeaaavksikaasstekfpdgfwlgeqlvcwaggttpwni 333
 Qy 241 FPVISLYLMGEVTNOSFRITILPQQYLPRPVEDVATSDQDCYKFAISQSSTGVMGAVIME 300
 Db 334 fpvislylmgevtngsfritilpqyylprpvedvatsqddcykfaissqstgtvmgavime 393
 Qy 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPFCHL 338
 Db 394 gfyvvdfrarkrigfavsachvhdefrtaavegpftvl 431

RESULT 7
 Y88432
 ID Y88432 standard; Protein; 459 AA.
 XX
 AC Y88432;
 XX
 DT 03-AUG-2000 (first entry)
 XX
 DE T7-caspase-human-pro-Asp-2(a)-deltaTM amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site;
 KW T7-caspase-human-pro-Asp-2(a)-deltaTM.

OS Chimeric - Homo sapiens.
 OS Chimeric - Bacteriophage t7.
 XX
 PN WO200017369-A2.
 XX
 PD 30-MAR-2000.
 XX
 PF 23-SEP-1999; 99WO-US20881.
 XX
 PR 24-SEP-1998; 98US-0101594.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX

PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 DR WPI; 2000-303209/26.
 DR N-PSDB; A15669.
 XX
 PT New enzyme designated human aspartase useful in research into
 PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 PT the beta secretase site to produce amyloid beta peptide -
 XX
 PS Example 9; Fig 7; 183pp; English.
 XX
 CC This sequence represents a modified version of the human aspartase 2
 CC (Asp2) nucleotide sequence. The sequence is used in the bacterial
 CC expression of human Asp2L. The invention relates to a protease
 CC (e.g. Asp2) capable of cleaving the beta secretase site of amyloid
 CC precursor protein (APP). The protease contains a sequence encoding the
 CC amino acid sequence DTG and a sequence encoding DSG or DTG separated by
 CC 100-300 amino acids. When mutated the APP gene causes an autosomal
 CC dominant form of Alzheimer's disease. APP localises to the cell surface
 CC membrane and have a single C-terminal transmembrane domain. Proteolytic
 CC processing of APP produces the amyloid beta protein, which is possibly
 CC very important in Alzheimer's disease. The invention includes a
 CC nucleotide sequence encoding the protease, a vector containing the
 CC screening for inhibitors of beta secretase activity are also given in the
 CC invention. The human aspartase protein and nucleotide sequences and the
 CC methods for identifying inhibitors of the protease, are useful in the
 CC treatment of and research in to Alzheimer's disease.
 XX
 SQ Sequence 459 AA;

Query Match 88.8%; Score 1668; DB 21; Length 459;
 Best Local Similarity 95.0%; Pred. No. 1.3e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;
 Qy 3 PFI--YLOAHFTLCGWSSTYRDLRKGVYVPYTOGKWEGLGTLVSIPIHGNVTVRANI 60
 Db 113 pflhryyqrql-----sstyrdlrkgyvpytgkgwgelgtdlvsiphgpnvtvrani 166
 Qy 61 AAITESDRFFINGSNWEGILGLAYAEIARPDSPFPFDSLKQTHVPLNLSLQLCGAGF 120
 Db 167 aaatesdkffingsnwegilglayaeiarpdpslepfdslykqthvplnlslhlcgagf 226
 Qy 121 PLNQSEVLASVGGSMIIGGIDHSlyTGSlyWTPPIRREWYVEIIVRVEINGODLKMDCKE 180
 Db 227 plnqsevlasvggsmilggidhslytgslywtpirrewyveilvrveingodlkmdccke 286
 Qy 181 YNDKSIIVDSGTTNLRPKKVFEEAAVKSIAASSTKEKFPDGFGLGEQLVLCWQAGTTPWNI 240
 Db 287 ynydksivdsdgttnlrpkpkvfeaaavksikaasstekfpdgfwlgeqlvcwaggttpwni 346
 Qy 241 FPVISLYLMGEVTNOSFRITILPQQYLPRPVEDVATSDQDCYKFAISQSSTGVMGAVIME 300
 Db 347 fpvislylmgevtngsfritilpqyylprpvedvatsqddcykfaissqstgtvmgavime 406
 Qy 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGGPFCHL 338
 Db 407 gfyvvdfrarkrigfavsachvhdefrtaavegpftvl 444
 RESULT 8
 Y88425
 ID Y88425 standard; Protein; 501 AA.
 XX
 AC Y88425;
 XX
 DT 03-AUG-2000 (first entry)
 XX
 DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.
 XX
 KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;
 KW Alzheimer's disease; beta secretase site.

XX OS Homo sapiens.
 XX PN WO200017369-A2.
 XX PD 30-MAR-2000.
 XX PF 23-SEP-1999; 99WO-US20881.
 XX PR 24-SEP-1998; 98US-0101594.
 XX PA (PHAA) PHARMACIA & UPJOHN CO.
 XX PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;
 XX DR WPI; 2000-303209/26.
 XX DR N-PSDB; A15662.
 XX PT New enzyme designated human aspartase useful in research into
 XX PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at
 XX PT the beta secretase site to produce amyloid beta peptide
 XX PS Claim 48; Fig 2; 183pp; English.
 XX CC This sequence represents the human aspartyl protease 2 (Asp2) amino acid
 XX CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 XX CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 XX CC protease contains a sequence encoding the amino acid sequence DTG and a
 XX CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
 XX CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 XX CC disease. APP localises to the cell surface membrane and have a single
 XX CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 XX CC the amyloid beta protein, which is possibly very important in Alzheimer's
 XX CC disease. The invention includes a nucleotide sequence encoding the
 XX CC protease, a vector containing the nucleotide sequence, and a cell line
 XX CC comprising the vector. Methods for screening for inhibitors of beta
 XX CC secretase activity are also given in the invention. The human aspartase
 XX CC protein and nucleotide sequences and the methods for identifying
 XX CC inhibitors of the protease, are useful in the treatment of and research
 XX CC in to Alzheimer's disease.
 XX SQ Sequence 501 AA;

Query Match 88.8%; Score 1668; DB 21; Length 501;
 Best Local Similarity 95.0%; Pred. No. 1.5e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;

QY 3 PFI--YLOAHFTLCGWSSTYRDLRKGVYVPTQCKWEGELGTLVSTPHGPNVTVRANI 60
 ||: |
 DB 107 pfhryyqrql-----sstyrdlrgvyvptqgkwegelgtdivsphgpnvtvrani 160
 QY 61 AAITESDKFFINGSNWEGILGAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120
 ||: |
 DB 161 aaatesdkffingsnwegilglayaeiarpdslpffdsllvkqthvplnlsqlcgagf 220
 QY 121 PLNQSEVLASVGGSMIIGIDHSLYTSGSLWYTPIRREWYEVIIIVRVEINGDLKMDCKE 180
 ||: |
 DB 221 plnqsevlasvggsmiigidhslytsgslwycprrwyyevilvrveingdldkmdcke 280
 QY 181 YNYDKSIVDSGTTNLRPKKVFEEAAVKSIKAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
 ||: |
 DB 281 ynydksivdsgttnlrpkkveaavksikaasstekfpdgfwlgeqlvcwqagtppwni 340
 QY 241 FPVISLYLMGEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 ||: |
 DB 341 fpvislylmgevtngsfrtilpqylrpvedvatsddcykfaissqstgtvmgavime 400
 QY 301 GFYVVFDRARKRIGFAVSACHVHDEFRTAAVGPFCCHL 338
 ||: |
 DB 401 gfyvvfdrarkrigrfavsachvhdefrtaavgepfcvtl 438

RESULT 9
 W59807 W59807 standard; Protein; 501 AA.
 XX AC W59807;
 XX DT 26-OCT-1998 (first entry)
 XX DE Amino acid sequence of human ASP2 (aspartic protease 2).
 XX KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
 XX KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
 XX KW prohormone processing.
 XX OS Homo sapiens.
 XX PN EP855444-A2.
 XX PD 29-JUL-1998.
 XX PF 27-JAN-1998; 98EP-0300573.
 XX PR 28-JAN-1997; 97GB-0001684.
 XX PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX PI Chapman CG, Murphy K, Powell DJ, Smith TS;
 XX DR WPI; 1998-389809/34.
 XX DR N-PSDB; V41696.
 XX PT New nucleic acid encoding human aspartic protease 2 - used to treat,
 XX PT prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
 XX PT processing
 XX PS Claim 1; Page 7; 26pp; English.
 XX CC This is the amino acid sequence of the human ASP2 (aspartic protease
 XX CC family), used in the method of the invention. Agonists and
 XX CC antagonists for ASP2 immunospecific antibodies are used to treat
 XX CC conditions requiring increased or decreased activity or expression of
 XX CC ASP2 respectively. ASP2 is used to treat and diagnose e.g.
 XX CC Alzheimer's disease, cancer and prohormone processing and ASP2 or a
 XX CC fragment can be used to induce an immune response against the above
 XX CC conditions.
 XX SQ Sequence 501 AA;

Query Match 88.8%; Score 1667; DB 19; Length 501;
 Best Local Similarity 95.0%; Pred. No. 2e-173;
 Matches 321; Conservative 1; Mismatches 8; Indels 8; Gaps 2;

QY 3 PFI--YLOAHFTLCGWSSTYRDLRKGVYVPTQCKWEGELGTLVSTPHGPNVTVRANI 60
 ||: |
 DB 107 pfhryyqrql-----sstyrdlrgvyvptqgkwegelgtdivsphgpnvtvrani 160
 QY 61 AAITESDKFFINGSNWEGILGAYAEIARPDSPPEFFDSLKVQTHVPLNLSQLCGAGF 120
 ||: |
 DB 161 aaatesdkffingsnwegilglayaeiarpdslpffdsllvkqthvplnlsqlcgagf 220
 QY 121 PLNQSEVLASVGGSMIIGIDHSLYTSGSLWYTPIRREWYEVIIIVRVEINGDLKMDCKE 180
 ||: |
 DB 221 plnqsevlasvggsmiigidhslytsgslwycprrwyyevilvrveingdldkmdcke 280
 QY 181 YNYDKSIVDSGTTNLRPKKVFEEAAVKSIKAASSTKEKPDGFWLGEQLVCWQAGTTPWNI 240
 ||: |
 DB 281 ynydksivdsgttnlrpkkveaavksikaasstekfpdgfwlgeqlvcwqagtppwni 340
 QY 241 FPVISLYLMGEVTNQSFRITILPQOYLRPVEDVATSDDCYKFAISQSSTGTVMGAVIME 300
 ||: |
 DB 341 fpvislylmgevtngsfrtilpqylrpvedvatsddcykfaissqstgtvmgavime 400

xx
SQ Sequence 476 AA;

Query Match 81.28; Score 1525.5; DB 21; Length 476;

Best Local Similarity 88.28; Pred. No. 5.2e-158;

Matches 298; Conservative 1; Mismatches 6; Indels 33; Gaps 3;

Qy 3 PFI--YLQAHFTLCGWSSTYRDLRKGVVYPTQGWEGELGTDLVSIHPGPNVTVRANI 60

Db 107 pfhryyqrql-----sstyrdlrkgyvpytgkwegelgtldvsihpognvtvrani 160

Qy 61 AAITESDKFFINGSNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVNPVLFSLQCGAGF 120

Db 161 aaitesdkffingsnwegilglayaeiar-----lcsagaf 195

Qy 121 PLNQSEVLASVGGSMIGGIDHSLYTGSWLTPTIRREHYEVIIVRVEINGQDLKMDCKE 180

Db 196 plnqsevlasvggsmilggidhslytgslytptirrehyeviivrvveingdldkmdcke 255

Qy 181 YNYSKSIYDVGTTNLRPKKVFEEAAVKSIAASTSEKPPDGFVWLGEOQLVCWQAGTTPWNI 240

Db 256 ynydksiydvgttnlrlpkkvfeavksikaasstekefpdgfwlgEOQLVCWQAGTTPWNI 315

Qy 241 FPIVSLYLMGEVTVNQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSGTGTVMGAVIME 300

Db 316 fpvislylmgevtngsfritilpqylrpvedvatsqdcykfaalsqsgstgtvmgavime 375

Qy 301 GFYVFDPRKRIGFAVSACHVHDEFRTAAVEGPFCHL 338

Db 376 gfyvfdprkrigfavsachvhdefrtaavegpfvtl 413

RESULT 13

W61362
ID W61362 standard; Protein; 518 AA.

xx W61362;

DT 25-SEP-1998 (first entry)

DE Aspartic proteinase ASPI.

KW ASPI; Aspartic proteinase; Alzheimer's disease; cancer; melanoma.

OS Homo sapiens.

xx EP848062-A2.

PN 17-JUN-1998.

PD 01-DEC-1997; 97EP-0309648.

PF 14-DEC-1996; 96GB-0026022.

PR (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Chapman CG, Evans JR, Powell DJ, Southan C;

DR WPI; 1998-314477/28.

DR N-PSDB; V27962.

xx New isolated polynucleotide encodes Aspartic protease polypeptide -

PT used to diagnosis, treat and vaccinate against Alzheimer's disease,

PT cancer and melanoma

xx Claim 11; Page 7; 19pp; English.

xx The human ASPI protein is structurally related to other proteins of the

CC Aspartic proteinase family. ASP 1 polypeptides and polynucleotides can

CC be used to diagnosis, treat and vaccinate against Alzheimer's disease,

CC cancer and melanoma.

xx
SQ Sequence 518 AA;

Query Match 49.38; Score 925; DB 19; Length 518;

Best Local Similarity 52.38; Pred. No. 2.5e-92;

Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3;

Qy 3 PFYIQAHTLCSGWSSTYRDLRKGVVYPTQGWEGELGTDLVSIHPGPNVTVRANIAA 62

Db 122 phsyldtyfd--tersstyrskgfdvtvkytgsgvtgfvgedlvtipkgnfsvlniat 179

Qy 63 ITESDKFFINGSNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVNPVLFSLQCGAGPPL 122

Db 180 ifesenffilgknwngilglayatlakpssslstffdsilvtqanlpnvfsmqmcgaglpv 239

Qy 123 NQSEVLASVGGSMIGGIDHSLYTGSWLTPTIRREHYEVIIVRVEINGQDLKMDCKEYN 182

Db 240 ags---gtngsgslvlggielpslykgdiwytpikeewyvgieiklieggqslnldcreyn 296

Qy 183 YDKSIVDVGTTNLRPKKVFEEAAVKSIAASTSEKPPDGFVWLGEOQLVCWQAGTTPWNIFF 242

Db 297 adkalvdsgttllrlpkkvfdavveavarslipedsdgtfwtgslactwnsetpwsyfp 356

Qy 243 VISLYLMGEVTVNQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSGTGTVMGAVIMEGF 302

Db 357 kisiylrdensrsfritilpqylqpmmgaglny-ecyrfgispsstnalvigatvmegf 415

Qy 303 YVFDPRKRIGFAVSACHVHDEFRTAAVEGPF 335

Db 416 yvifdradkrigfavsacaelagaaavseisgpf 448

RESULT 14

Y41714
ID Y41714 standard; Protein; 518 AA.

xx Y41714;

DT 07-DEC-1999 (first entry)

DE Human PR0852 protein sequence.

KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;

KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;

KW secreted protein; transmembrane protein.

OS Homo sapiens.

xx WO9946281-A2.

PD 16-SEP-1999.

PF 08-MAR-1999; 99WO-US05028.

PR 10-MAR-1998; 98US-0077450.

PR 11-MAR-1998; 98US-0077632.

PR 11-MAR-1998; 98US-0077641.

PR 11-MAR-1998; 98US-0077649.

PR 12-MAR-1998; 98US-0077791.

PR 13-MAR-1998; 98US-0078004.

PR 17-MAR-1998; 98US-0040220.

PR 20-MAR-1998; 98US-0078886.

PR 20-MAR-1998; 98US-0078910.

PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.

PR 25-MAR-1998; 98US-0079294.

PR 26-MAR-1998; 98US-0079656.

PR 27-MAR-1998; 98US-0079663.

PR 27-MAR-1998; 98US-0079664.

PR 27-MAR-1998; 98US-0079689.

PR 27-MAR-1998; 98US-0079728.

PR 27-MAR-1998; 98US-0079786.


```

XX 24-DEC-1998; 98WO-US27608.
XX PF
XX 31-DEC-1997; 97US-0070112.
XX PR
XX (CHIR ) CHIRON CORP.
XX PA
XX XX
XX Gliese K, Xin H;
XX PI
XX WPI; 1999-430248/36.
XX DR
XX N-PSDB; X84708.
XX DR
XX New polynucleotides associated with cancer metastasis
XX PT
XX Claim 4; Page 78-80; 80pp; English.
XX PS
XX This sequence represents a polypeptide of the invention, and is
XX CC an aspartyl-type protease, designated CSP56. The polynucleotides (PNS) of
XX CC the invention encode metastatic marker protein variants. The PNS and
XX CC polypeptides can be used as markers for cancer metastasis. The PNS can
XX CC be used for identifying metastatic tissue or metastatic potential of
XX CC a tissue, e.g. breast or colon tissue. They can also be used for
XX CC screening test compounds for the ability to suppress the metastatic
XX CC potential of a tumour. The products can be used for developing products
XX CC for the therapy of cancers, particularly breast or colon cancer.
XX CC
XX SQ Sequence 518 AA;

Query Match 49.3%; Score 925; DB 20; Length 518;
Best Local Similarity 52.3%; Pred. No. 2.5e-92;
Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3

QY 3 PFYIQLAHFTLCGWSSTYRDLRKGVVYPYTGKGWEGELGTDLVSIPHGPNVTVRANIAA 62
DB 122 phsyidytdf--terstyrskgfdvtykycgswtgrfvgedlvtipkgnfscflvniat 179
QY 63 ITESDKFFINGSNWEGILGLAYAEIARDDSPPEPFDLSLVKQTHVPNLVSLQCGAGFPL 122
DB 180 ifesenfflpkgilgwngilglavatliakpsssletffdsilvtqanlpnvfsmcagaglpv 239
QY 123 NQSEVLASVSGSMIIGGDHSLYGLSWYTPIRRWYVEYEVIVRVEINGQDLKMDCKEYN 182
DB 240 ags---gtngsvllvgliepslykgdlwytpikeesvyygieilkieigqgslnldcreyn 296
QY 183 YDKSIVDSGTTNLRLPKPVFAAVKSIKAASSTKFPFGFWLGEOLVCWQAGTTPWNIPF 242
DB 297 adkaivdsqgtllrlpqkfvdaavearaslipfsgdfwtgsglactwnsetpwsyfp 356
QY 243 VISLYLGEVNTNQFRITILPQOYLRPVEDVATSDDCYKFAISQSSGTGTVMGAVIMEGF 302
DB 357 kislyldensrsfritilpqylqipmmgaglny-ecyrfgispsstnalvigatvmegf 415
QY 303 YVFDRAKRTGFVAVSACHVHDEFPTAAVEGPF 335
DB 416 yvifdrakrvyfaaspcaeiaagavseisgpf 448

```

Search completed: March 24, 2001, 13:22:25
Job time: 56 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:34 ; Search time 73.46 Seconds
(without alignments)
560.032 Million cell updates/sec

Title: US-09-215-435-167
Perfect score: 1878
Sequence: 1 MVPEYILQAHFTLCGWSST.....EGPFCHLGRLWLQHSSTR 351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_15:*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertebrate.*
- 14: sp.unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1745	92.9	396	4 Q9ULS1	Q9uls1 homo sapien
2	913	48.6	514	11 Q9JL18	Q9jl18 mus musculus
3	784.5	41.8	213	4 Q9P0D2	Q9p0d2 homo sapien
4	767.5	40.9	396	4 Q9NZL1	Q9nzl1 homo sapien
5	726	38.7	468	4 Q9NZL2	Q9nzl2 homo sapien
6	559.5	29.8	255	11 Q9R1P7	Q9rlp7 mus musculus
7	254	13.5	423	5 Q9VKP7	Q9vkp7 drosophila
8	253.5	13.5	382	13 Q9PRG9	Q9prg9 gallus gall
9	253.5	13.5	391	5 Q9VRP6	Q9vrp6 drosophila
10	237.5	12.6	372	5 Q9VLK3	Q9vln3 drosophila
11	235	12.5	380	6 Q28950	Q28950 sus scrofa
12	233	12.4	396	13 Q93428	Q93428 chionodraco
13	232	12.4	346	6 Q9TS27	Q9ts27 bos taurus
14	232	12.4	384	13 Q91322	Q91322 rana catesb
15	230.5	12.3	385	6 Q29080	Q29080 sus scrofa
16	226.5	12.1	376	13 Q9PUR8	Q9pur8 pleuronecte
17	225.5	12.0	387	6 Q9N2D4	Q9n2d4 callithrix
18	225	12.0	381	6 Q9N2D2	Q9n2d2 callithrix
19	221	11.8	379	11 Q9JUX1	Q9jjx1 rattus norv

20	220.5	11.7	378	13 Q9PUR9	Q9pur9 pleuronecte
21	219	11.7	387	11 Q9JKE6	Q9jke6 mus musculus
22	216	11.5	388	6 Q46524	Q46524 felis silve
23	215	11.4	399	13 Q93458	Q93458 podarcis si
24	212.5	11.3	380	6 Q9TTX0	Q9ttx0 capra hircu
25	210.5	11.2	345	6 Q27951	Q27951 bos taurus
26	210.5	11.2	388	6 Q46523	Q46523 equus zebra
27	210.5	11.2	444	5 Q21966	Q21966 caenorhabdi
28	210	11.2	366	6 Q9N1P5	Q9n1p5 bubalus bub
29	210	11.2	387	11 Q9JUX2	Q9jjx2 rattus norv
30	210	11.2	398	13 P87370	P87370 oncorhynch
31	209	11.1	388	6 Q9N2D3	Q9n2d3 callithrix
32	208	11.1	446	5 Q9N9H3	Q9n9h3 necator ame
33	206	11.0	389	13 Q9W643	Q9w643 gallus gall
34	206	11.0	389	13 Q9PK1	Q9pk1 gallus gall
35	204	10.9	413	3 Q14413	Q14413 pichia angu
36	203	10.8	375	6 Q46500	Q46500 bos taurus
37	202.5	10.8	410	5 Q9VQ13	Q9vq13 drosophila
38	201	10.7	632	10 Q9SN13	Q9sn13 arabidopsis
39	200	10.6	383	5 Q76856	Q76856 dictyostell
40	199	10.6	370	6 Q9TTW1	Q9ttl1 bos taurus
41	198	10.5	389	6 Q9MYK3	Q9myk3 sus scrofa
42	197.5	10.5	376	6 Q46499	Q46499 bos taurus
43	197	10.5	389	6 Q9MYK2	Q9myk2 sus scrofa
44	196.5	10.5	450	5 Q76965	Q76965 plasmodium
45	196	10.4	395	5 Q9VEK3	Q9vek3 drosophila

ALIGNMENTS

RESULT 1					
Q9ULS1					
ID Q9ULS1	PRELIMINARY;	PRT;	396 AA.		
AC Q9ULS1					
DT 01-MAY-2000 (Tremblrel. 13, Created)					
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)					
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)					
DE KIAA1149 PROTEIN (FRAGMENT).					
GN KIAA1149.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=BRAIN;					
RA Hiroseawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;					
RT "Characterization of cDNA clones selected by the GeneMark analysis					
RL from size-fractionated cDNA libraries from human brain.";					
RL DNA Res. 6:329-336(1999).					
DR EMBL; AB032975; BAA86463.1;					
DR HSSP; P56272; IAMS.					
DR INTERPRO; IPR001461;					
DR PFAM; PF00026; asp; 2.					
DR PRINTS; PR00792; PEPSIN.					
FT NON_TER 1 1					
SQ SEQUENCE 396 AA; 44428 MW; A2CBDCD52DCC089E0 CRC64;					
Query Match 92.9%; Score 1745; DB 4; Length 396;					
Best Local Similarity 99.1%; Pred. No. 5.9e-145;					
Matches 330; Conservative 0; Mismatches 3; Indels 0; Gaps 0;					
Qy 6 YLQAHFTLCGWSSTYRDLRKGVVPTQGWEGELGTDLVSIHPGPNVTVRANIAATE 65					
Db 1 YLQAHFTLCGWSSTYRDLRKGVVPTQGWEGELGTDLVSIHPGPNVTVRANIAATE 60					
Qy 66 SDKFFINGSNWEGILGLAYAEIARPDSPFPFSLVKQTHVNPVLFSLQLCGAGFPLNQ 125					
Db 61 SDKFFINGSNWEGILGLAYAEIARPDSPFPFSLVKQTHVNPVLFSLQLCGAGFPLNQ 120					
Qy 126 EVLASVGGSMIIIGDHSGLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKREYN 185					

Db 121 EVLASVGGSMIGGIDHSLYTSGLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNDK 180
QY 186 SIVDSGTTNLRPKKVFEEAAVSKAASSTKPPDGFGLVQWQAGTTPWNIFFPVIS 245
Db 181 SIVDSGTTNLRPKKVFEEAAVSKAASSTKPPDGFGLVQWQAGTTPWNIFFPVIS 240
QY 246 LYLMEVNTQSPRITLPOQYLRPEVDVATSDQDCYKFAISQSGTGTVMGAVIMEGFYV 305
Db 241 LYLMEVNTQSPRITLPOQYLRPEVDVATSDQDCYKFAISQSGTGTVMGAVIMEGFYV 300
QY 306 FDRARRKRGFAVSACHVHDEFRTAAVEGPFCHL 338
Db 301 FDRARRKRGFAVSACHVHDEFRTAAVEGPFVIL 333
RESULT 2
Q9JL18 PRELIMINARY; PRT; 514 AA.
AC Q9JL18;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Choi D.K., Sugano S., Sakaki Y.;
RT "Molecular characterization of the mouse Aspl gene, a homolog of the
human ASP1 (Down Syndrome Region aspartyl protease).";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF216310; AAF36599.1; -;
KW Protease.
SQ SEQUENCE 514 AA; 55799 MW; A70725P2C1DF5B47 CRC64;

Query Match 48.6%; Score 913; DB 11; Length 514;
Best Local Similarity 52.0%; Pred. No. 8.5e-72;
Matches 173; Conservative 55; Mismatches 99; Indels 6; Gaps 3;
QY 3 PFYIQAHTFLCSGWSSTYRDLRKGVVVPYTGKWEGLGTLVSLPHGPNVTVRANIAA 62
Db 118 PHSYIDYFD--SESSSYHSGFDVTYKTSQSWTGFVGEDLVTPKGFNSFLVNIAT 175
QY 63 ITESDKFFINGNWEGILGLAYAEIARPDSPPEFDSLVKQTHVFNLSQLCGAGFPL 122
Db 176 IFESENFPLPGIKWNGILGLAYAAALAKPSSSLETFDLSLVAQAKIPDIIFSMQCGAGLPV 235
QY 123 NQSEVLASVGSMTIGIDHSLYTSGLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYN 182
Db 236 AGS---GTNGSGLVIGIEPSLYKGDILWYTPKEEYVQIEILKLEIGONLMDCREYN 292
QY 183 YKSDIVDSGTTNLRPKKVFEEAAVSKAASSTKPPDGFGLVQWQAGTTPWNIFF 242
Db 293 ADRAIVDSGTTNLRPKQKVFEEAAVARTSLIPESDGFMTGAQLACWTNSETPWAYFP 352
QY 243 VISLYLMEVNTQSPRITLPOQYLRPEVDVATSDQDCYKFAISQSGTGTVMGAVIMEGF 302
Db 353 KISYLRDENASRFRITLPOQYLRPEVDVATSDQDCYKFAISQSGTGTVMGAVIMEGF 411
QY 303 YVVEDRARRKRGFAVSACHVHDEFRTAAVEGPF 335
Db 412 YVVEDRARRKRGFAVSACHVHDEFRTAAVEGPF 444
RESULT 3
Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE HSPC104 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF161367; AAF28927.1; -;
FT NON_TER
SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

Query Match 41.8%; Score 784.5; DB 4; Length 213;
Best Local Similarity 66.0%; Pred. No. 4.4e-61;
Matches 157; Conservative 8; Mismatches 24; Indels 49; Gaps 4;
QY 138 GGIDHSLYTSGLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTGTTNLR 197
Db 1 GGDHSLYTSGLWTPYIRREWYEVIIIVRVEINGQDLKMDCKEYNDKSIYDSTGTTNLR 60
QY 198 PKKVFEEAAVSKAASSTKPPDGFGLVQWQAGTTPWNIFFVISLYLMEVNTQSF 257
Db 61 PKKVFEEAAVSKAASSTKPPDGFGLVQWQAGTTPWNIFFVISLYLMEVNTQSF 120
QY 258 RTITLPOQYLRPEVDVATSDQDCYKFAISQSGTGTVMGAVIMEGFVYVDFRARRKRGFAV 317
Db 121 RTITLPOQYLRP-----WKMPRPKTTVTVCHLTVI-----GHY 156
QY 318 SACHVHDEFRTAAVEGP-----FCHLGHGRMLQHSR 351
Db 157 GSCY-HGGLLPLSLGIPENELACLSALACTMSSGRQKALCHLGHGRMLQHSR 213
RESULT 4
Q9NZL1 PRELIMINARY; PRT; 396 AA.
AC Q9NZL1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188277; AAF35836.1; -;
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match 40.9%; Score 767.5; DB 4; Length 396;
Best Local Similarity 52.4%; Pred. No. 3.4e-59;
Matches 150; Conservative 46; Mismatches 75; Indels 15; Gaps 5;
QY 3 PFYIQAHTFLCSGWSSTYRDLRKGVVVPYTGKWEGLGTLVSLPHGPNVTVRANIAA 62
Db 122 PHSYIDYFD--TERSTYRSKGFDTYKTSQSWTGFVGEDLVTPKGFNSFLVNIAT 179
QY 63 ITESDKFFINGNWEGILGLAYAEIARPDSPPEFDSLVKQTHVFNLSQLCGAGFPL 122

Db 180 IFESNFPLPGIKWNGILGLAYATLAKPSSSLETFDLSVLTQANIPNVFQMCGAGLPV 239
QY 123 NQSEVLASVGGSMIIIGDHSLYTGLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYN 182
Db 240 AGS---GTNGSLVGGIEFSLYKGDWYPIKEEWYQIEILKEITGGQSLNLDCKEYN 296
QY 183 YKSDIVSGTTLNLRPKKVFEEAAKASSTKFFDPGFWLGEOLVCWQAGTTPWNIFF 242
Db 297 ADKAIVDSGTTLLRLPKQKFDVAVAVARASLIPEFSDGFWGSQLACWTNSETPWSYFP 356
QY 243 VISIYLMGEVNTQSFRTITLPOQYLRVEDVATSDDCYKF-ALISQ 287
Db 357 KISIYLRDENSSRSFRITILPQK-LRVLQ-----CLKPPGLSQ 393

RESULT 5

Q9NZL2 ID Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC Q9NZL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Solans A., Estivill X., de la Luna S.;
RT "Cloning of a novel mammalian aspartyl protease";
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188276; AAF35835.1;
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 38.7%; Score 726; DB 4; Length 468;
Best Local Similarity 43.5%; Pred. No. 1.9e-55;
Matches 145; Conservative 50; Mismatches 82; Indels 56; Gaps 4;

QY 3 PFYIQAHTFLCSGWSSTYRDLRGVVPYVYQGWEGELGDLVSIPIHGPNTVVRANIAA 62
Db 122 PHSYDIYFD--TERSTYSRKGFVDVYKYTGSGTGFVGEDLVTPKGFNTSELVNIAT 179
QY 63 ITESDKFFINGSNWEIGLGLAYAIARPDSPPEFFDSLVKQTHVPLNLSQLCGAGPL 122
Db 180 IFESNFPLPGIKWNGILGLAYATLAKPSSSLETFDLSVLTQANIPNVFQMCGAGLPV 239
QY 123 NQSEVLASVGGSMIIIGDHSLYTGLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYN 182
Db 240 AGS---GTNGSLVGGIEFSLYKGDWYPIKEEWYQIEILKEITGGQSLNLDCKEYN 296
QY 183 YKSDIVSGTTLNLRPKKVFEEAAKASSTKFFDPGFWLGEOLVCWQAGTTPWNIFF 242
Db 297 ADKAIVDSGTTLLRLPKQKFDVAVAVARASLIPEFSDGFWGSQLACWTNSETPWSYFP 356
QY 243 VISIYLMGEVNTQSFRTITLPOQYLRVEDVATSDDCYKF-ALISQ 287
Db 330 -----YIQPMGAGLNY-ECYRFGISPSSTNALVIGATVMEGF 365
QY 303 YVTFDRARKRIGFAVSACHVIDEFRTAAVEGPF 335
Db 366 YVTFDRARKRIGFAVSACHVIDEFRTAAVEGPF 335
QY 366 YVTFDRARKRIGFAVSACHVIDEFRTAAVEGPF 335
Db 366 YVTFDRARKRIGFAVSACHVIDEFRTAAVEGPF 335

RESULT 6

Q9R1P7 ID Q9R1P7 PRELIMINARY; PRT; 255 AA.
AC Q9R1P7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ASPARTYL PROTEASE (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;
RT "Cloning of a gene from chromosome 21 Down Region encoding a potential transmembrane protease";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051150; AAD45964.1;
DR INTERPRO; IPR001461;
DR INTERPRO; IPR001969;
DR PFAM; PF00026; asp; 2;
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 29.8%; Score 559.5; DB 11; Length 255;
Best Local Similarity 53.8%; Pred. No. 3.1e-41;
Matches 100; Conservative 34; Mismatches 51; Indels 1; Gaps 1;

QY 150 WYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNKSDVSGTTLNLRPKKVFEEAAKASI 209
Db 1 WYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNKSDVSGTTLNLRPKKVFEEAAKASI 209
QY 210 KAASSTKFFDPGFWLGEOLVCWQAGTTPWNIFFVVISLVMGEVNTNOSFRITLPQYLRP 269
Db 61 ARTSLIPEFSDGFWTGAQLACWTNSETPWYFPKISIVLRDENASRSFRITLPQYLRP 120
QY 270 VEDVATSDDCYKFAISOSSTGTVMGAVIMEGFYVVDPRARKRIGFAVSACHVIDEFRTA 329
Db 121 MNGAGFNY-ECYRFGISSTNALVIGATVMEGFYVVDPRARKRIGFAVSACHVIDEFRTA 329
QY 330 AVEGPF 335
Db 180 EISGPF 185

RESULT 7

Q9VRP7 ID Q9VRP7 PRELIMINARY; PRT; 423 AA.
AC Q9VRP7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG6508 PROTEIN.
GN CG6508.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reineb K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski R.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003630; AAF53015.1; -.
DR HSSP: P07339; ILIYB.
DR FLYBASE: FBgn0032303; CG6508.
DR INTERPRO: IPR001461; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00026; asp; 1.
DR PRINTS: PR00792; PEPSTN.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
DR SEQUENCE 423 AA; 46509 MW; 0195CA016EAA5C2C CRC64;

Query Match 13.5%; Score 254; DB 5; Length 423;
Best Local Similarity 28.4%; Pred. No. 4.2e-14;
Matches 93; Conservative 56; Mismatches 132; Indels 46; Gaps 13;

QY 8 QAHTLCGWSSTYRLRGVYVYTGKWEGLGTLVSIHPGNVTVRANAAITESD 67
DB 113 QKHKNYSSASSSHVEDGKFSIQYSGSLGFLSTDTVDI-DGMVIRNQTAEAIDEPG 171
QY 68 KFFINGNWGILGLAYAEATARDPDPPEFDSLVKQTHVFN-LFSIQLCGAGFPLNQS 126
DB 172 SAFVN-TIFDGIIGMAFASISGGVTT-DFNIIQGLVKHPVFSVYLRDQ--TSQS- 224
QY 127 VLASVGSMIIGGDHSLYTGSLWYTPIRREWYEVYLIIVRVEINGDLMKDCKEYNDKS 186
DB 225 -----GGEVWGIDRSIYRGCIYVPSVMPAQWQTANSVKIEGILLNCG-----QA 273
QY 187 IVDSGTTNLRPKKVFEEAAVKSIAK-----ASSTKFPDGVWGLQVQAGTTPWTFPVI 244
DB 274 IADGTSLIAVPLRAYKAIKNVLNATDAGDEAFVDCSSLCR-----LPNV 319
QY 245 SLVLMGEVNTQSFRTILPQOYLRPVEDVATSQDDCYK-PAISQSSTGTVMGAVIMEGEY 303
DB 320 NLNIGGTT-----YTUPKDYIKVQ-ADNQTCLSGFTYLGQNLNLWILGIFLGKVY 372
QY 304 VFEDRARKRIGFAVSACHVDFEPTAA 330
DB 373 TVFDVGKERIGFA--KLKKHSYRYVA 397

RESULT 8
Q9PRG9 PRELIMINARY; PRT; 382 AA.
AC Q9PRG9;
OC Pterygota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE PEPSINOGEN A.
GN PCGA.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakamoto N., Yasugi S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Sakamoto N., Saiga H., Yasugi S.;
RT "Analysis of temporal expression pattern and cis-regulatory sequences
of chicken pepsinogen A and C.";
RL Biochem. Biophys. Res. Commun. 250:420-424(1998).
DR EMBL: AB025283; BAA77268.1; -.
DR EMBL: AB025281; BAA76891.1; -.
DR HSSP: P00794; 4CMS.
DR INTERPRO: IPR001461; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00026; asp; 1.
DR PRINTS: PR00792; PEPSTN.
DR PROSITE: PS00141; ASP_PROTEASE; 2.
DR SEQUENCE 382 AA; 42056 MW; C3C7EB8CA4DD3E28 CRC64;

Query Match 13.5%; Score 253.5; DB 13; Length 382;
Best Local Similarity 23.6%; Pred. No. 4e-14;
Matches 80; Conservative 62; Mismatches 114; Indels 83; Gaps 13;

QY 2 VPFIYLOAHFTLCGWSSTYRLRGVYVYTGKWEGLGTLVSIHPGNV 54
DB 100 VPSIYCKS--SACSNHKKRFPDPSKSTYVSTNETVIAYGTGSMGILGYDTAV--SSI 154
QY 55 TVRANAAITESDK-FFINGSNWEGILGLAYAEIARPDSPPEFDSLVKQTHV-PNLFS 112
DB 155 DVQNOIFGLSETEPGSFYFCNFDFGILGLAFPSIS--SSGATPVFDNMMSOHLVAQDLFS 212
QY 113 LQLCGAGFPLNQLSEVLASVGGSMIIGGDHSLYTGSLWYTPIRREWYEVYLIIVRVEINGQ 172
DB 213 VYLSKDG-----ETGSGVFLGGIDPNYTTIGIYWVPLSAETIWOITMDRVTVGK 262
QY 173 DLK--MDCKEYNDKSIVDSGTTNLRPKKVFEEAAVKSIAKASSTE-----KFPDGEW 223
DB 263 YVACFFTC-----QAIVDTGTSLLVMPQAGYRNRIKDLGVSSDGEISCDISKLPD--- 313
QY 224 LGELVQWQAGTTPWNIFFPVISLYLMGEVNTQSFRTILPQOYLRPVEDVATSQDDCYK 283
DB 314 -----VTFHNGHA-----FTLPASAYVLNEDGSCML 340
QY 284 AISOSSSTGT-----VMGAVIMEGEYVVFDRARKRIGFA 316
DB 341 GFENMGTPPELGEQWILGDVFIREFYVIFDRANKVGLS 379

RESULT 9
Q9VKP6 PRELIMINARY; PRT; 391 AA.
AC Q9VKP6;
OC Pterygota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```



```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL Science 287:2185-2195(2000).
DR EMBL; AE003621; AAF52686.1;
DR HSSP; P00797; 2REN.
DR FLYBASE; FBgn0032049; CG13095.
DR INTERPRO; IPR001461;
DR INTERPRO; IPR001969;
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE;
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN 2.
SQ SEQUENCE 372 AA; 40080 MW; D45469E8AD72FCCE CRC64;
```

```
Query Match 13.5%; Score 253.5; DB 5; Length 391;
Best Local Similarity 28.2%; Pred. No. 4.1e-14;
Matches 88; Conservative 54; Mismatches 127; Indels 43; Gaps 12;

Qy 8 QAHFTLCGNSSTYDLRGKGVVYPTQKWECEGLTDLVSTPHGPNVTVRANIAITSD 67
Db 115 QRKNKYDSSASTYVANGEEFAIEYGTGSLGSLNDIVTIA-GISTQNTFGALSEPG 173
Qy 68 KFFINGSNWEIGLGLAYAEIARPDSPFPFFDSLVKQTHVFN-LFSLQLCAGGFLNQSE 126
Db 174 TTFVD-APFAGILGLAFSAIA--VDGVTPFPDNNISQGLDEPVISFYLLKRGQ----- 223
Qy 127 VLASVGSMTIGGDHSLYLTGSLMTPYIRREWYVEYIVRVEINGQDLKMDKEYNDKS 186
Db 224 -TAVRGGEILGGIDSSLYRSLTVVPVPSVPAYWQVKVNTKTNGTLLCNGC-----QA 276
Qy 187 IVDSTTNLRPLPKKVFEEAAVKSIAASTKPEPDGFWLGEQLV-CWQAGTTPWNIFFVIS 245
Db 277 IADTGTSLIAVPLAAYRKINQLGATDND-----GEAFVRCGRVSS-----LPRKV 322
Qy 246 LYLMEVNTNQSFRTILPQQLRPPVEDVATSDDCYK-FAISQSSTGTVMGAVIMEGVYV 304
```


ID Q29080 PRELIMINARY; PRT: 385 AA.
AC Q29080:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE PEPSINOGEN A PRECURSOR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89174702; PubMed=2494172;
RA Lin X.L., Wong R.N.S., Tang J.;
RT "Synthesis, purification,, and active site mutagenesis of recombinant
RT porcine pepsinogen.,";
RL J. Biol. Chem. 264:4482-4489(1989).
DR EMBL; J04601; AAA31096.1; -.
DR HSSP; P00791; IPSA.
DR INTERPRO; IPR001461; -.
DR INTERPRO; IPR001791; -.
DR INTERPRO; IPR001969; -.
DR PFAM; PF00026; asp; 1.
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Signal; Hydrolase; Aspartyl protease.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 385 POTENTIAL.
FT CHAIN 60 385 POTENTIAL.
SQ SEQUENCE 385 AA; 41310 MW; 9ABC40GAD5877708 CRC64;

Query Match 12.3%; Score 230.5; DB 6; Length 385;
Best Local Similarity 25.8%; Pred. No. 4.2e-12;
Matches 85; Conservative 60; Mismatches 125; Indels 59; Gaps 15;

QY 2 VPRTYLAQHFTLCGSGW-----SSTYRLRGVYVYPTQGWEGELGDLVSIHPGPNV 54
DB 99 VPSVYCSS--LACSDHNFQNPDDSTFEATSQELSITYGTGSMTGILGYDTYQV---GGI 153
QY 55 TVRANTAAITESDK-FFINGSNWEGILGLAYAEIARPDSPPEPFDLSLVKQHV-PNLFS 112
DB 154 SDTNQIFGLSETEPGFLYAPFDGILGLAYPSISA--SGATPVFDNLWDQGLVSDQLFS 211
QY 113 LOLCGAGFPLNQSEVLASVGSGLIGDHSYLTGSLWTPIRREWYEVIIIVRVEINGQ 172
DB 212 VYLSS-----NDDS-----GSVLLGGIDSSYTGSLNWVPVSVGEYQWITLDSITMDGE 261
QY 173 DLKMD--CKEYNDKSIDVSGTTLNLRPKVFEAAVKSIAASSTKFKPDPGFWLGEOLV- 229
DB 262 TIACSGGC-----QAIVDTGTSLLTGPTSAIANIQSDIGASENS-----YGEWVIS 307
QY 230 CWQAGITPNWIFPVISLYLMGEVNTNSFRITILPQOYLRPVEDVATSDDDCYKFAISQSS 289
DB 308 CSSIDSLPDIIVF-----TINGVQPLSPSAYIIQDDDDSCSTSG---FEGMDVPTS 353
QY 290 TGT--VNGAVIMEGFYVVFDRARKRIGFA 316
DB 354 SGELWILGDVFIROYTYVDFRANKVGLA 382

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2001, 13:21:30 ; Search time 44.26 Seconds
(without alignments)
538.480 Million cell updates/sec

Title: US-09-215-435-167

Perfect score: 1878

Sequence: 1 MVPFTYLOAHFTLCGWSST.....EGPFCHLGHRLQLQSTDR 351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:*

1: pirl:*

2: pirl2:*

3: pirl3:*

4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1673	89.1	501	2 A59090	aspartic proteinas
2	253.5	13.5	367	1 JECH	pepsin A (EC 3.4.2
3	253.5	13.5	382	2 JE0370	pepsin A (EC 3.4.2
4	251	13.4	388	1 S19682	pepsin A (EC 3.4.2
5	239	12.7	387	2 B38302	pepsin (EC 3.4.23
6	237.5	12.6	387	2 E38302	pepsin (EC 3.4.23
7	235	12.5	380	2 I47176	chymosin (EC 3.4.2
8	234.5	12.5	388	1 S19684	pepsin A (EC 3.4.2
9	234	12.5	386	1 PEPG	pepsin A (EC 3.4.2
10	233	12.4	398	2 S66465	cathepsin E (EC 3
11	232	12.4	384	2 A39314	gastricsin (EC 3.4
12	231.5	12.3	388	1 PERU	pepsin A (EC 3.4.2
13	231.5	12.3	388	1 PEMQAR	pepsin A (EC 3.4.2
14	230.5	12.3	383	2 A41443	pepsin (EC 3.4.23
15	230	12.2	387	2 C38302	pepsin (EC 3.4.23
16	230	12.2	396	2 A34401	cathepsin E (EC 3
17	229.5	12.2	388	1 PEMQAJ	pepsin A (EC 3.4.2
18	229.5	12.2	391	2 A43356	cathepsin E (EC 3
19	229	12.2	387	2 D38302	pepsin (EC 3.4.23
20	227.5	12.1	388	2 A30142	pepsin A (EC 3.4.2
21	227	12.1	381	1 CMSHB	chymosin (EC 3.4.2
22	225.5	12.0	387	2 JC7245	pepsinogen A - com
23	225.5	12.0	388	2 B30142	pepsin A (EC 3.4.2
24	225	12.0	381	2 JC7247	prochymosin - comm
25	224	11.9	412	1 KHHUD	cathepsin D (EC 3
26	220	11.7	396	2 S36865	cathepsin E (EC 3
27	219	11.7	381	1 CMBO	chymosin (EC 3.4.2
28	218	11.6	377	1 PEMOCJ	gastricsin (EC 3.4
29	218	11.6	389	2 A38302	pepsin (EC 3.4.23

30 216.5 11.5 407 1 KHRTD
31 215.5 11.5 334 2 JC4870
32 212 11.3 410 1 KHMSD
33 210.5 11.2 444 2 T24204
34 210 11.2 398 2 I51185
35 209 11.1 388 2 A29937
36 209 11.1 388 2 JC7246
37 207.5 11.0 365 2 S66466
38 206 11.0 344 1 KHPGD
39 206 11.0 389 2 JE0371
40 202 10.8 394 2 B43356
41 201 10.7 632 2 T45858
42 196 10.4 376 2 I45856
43 192.5 10.3 389 2 I46616
44 191.5 10.2 380 2 S03433
45 190 10.1 387 2 I46617

ALIGNMENTS

RESULT 1

A59090

aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N;Alternate names: beta-secretase; beta-site APP cleaving enzyme

C;Species: Homo sapiens (man)

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000

C;Accession: A59090

R;Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.; Tepl

M.A.; Biere, A.L.; Curran, E.; Burgess, T.; Louis, J.C.; Collins, F.; Treanor, J.; Ro

Science 286, 735-741, 1999

A;Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the tran

A;Reference number: A59090; MUID:20002972

A;Note: submitted to Genbank, September 1999

A;Accession: A59090

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-501 <VAS>

A;Cross-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539

C;Genetics:

A;Gene: BACE

C;Superfamily: beta-secretase

C;Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein; hydrolase;

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-45/Domain: propeptide #status predicted <PRO>

F;46-501/Product: acid proteinase BACE #status predicted <MAT>

F;461-477/Domain: transmembrane #status predicted <TRN>

F;93,289/Active site: Asp #status predicted

F;153,172,223,354/Binding site: carboxylate (Asn) (covalent) #status predicted

F;330-380/Disulfide bonds: #status predicted

Query Match 89.1%; Score 1673; DB 2; Length 501;
Best Local Similarity 95.3%; Pred. No. 4.2e-137;
Matches 322; Conservative 1; Mismatches 7; Indels 8; Gaps 2;

QY 3 PFI--YLOAHFTLCGWSSTYRDLKRGVYVPTQKGEGELGTDLVSLPHGPNVTVRANI 60

Db 107 PFLHRYTQRQL-----SSTYRDLKRGVYVPTQKGEGELGTDLVSLPHGPNVTVRANI 160

QY 61 AAITEBKFFINGNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVPLNLSLQCGAGF 120

Db 161 AAITEBKFFINGNWEGILGLAYAEIARPDSPPEPFDLSLVKQTHVPLNLSLQCGAGF 220

QY 121 PLNSEVLASVGGSMIIIGIDHSLYTGSLWYTPIRREWYVEIIVRVVEINGDLKMDCKE 180

Db 221 PLNSEVLASVGGSMIIIGIDHSLYTGSLWYTPIRREWYVEIIVRVVEINGDLKMDCKE 280

QY 181 YNYDKSIVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKFPDGFVLGEQLVCWQAGTTPWNI 240

Db 281 YNYDKSIVDSGTTNLRPLPKKVFEEAAVKSIIKAASSTKFPDGFVLGEQLVCWQAGTTPWNI 340

QY 241 FPIVLSILMGEVTNQSPRITILPOOYLPRPVEDVATSDQDCYKFAISOSSTCTVMGAVIME 300

[illegible]

Db 123 SSTYRATSKTVSITYGTGSMTGILGYDTVKV---GGISDNTQIFGLSETEPGFFLYPAPF 179
Qy 77 EGILGLAYAEIARDDSPPEFFDSLVKQTHV-PNLFSLQCGAGFPPLNQSEVLASVGGSM 135
Db 180 DGIILGLAYPSIS--SSGATPVFQINQRLVSODLFSVLSAD-----DQS-----GSVV 227
Qy 136 IIGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGQDL--KMDCKEYNDKSIIVDSGTT 193
Db 228 IFGGIDSSYTGSLNWPVSVGEYQWISVDSITMNGKTIACAKG-----QAIVDTGTS 281
Qy 194 NLRLPKKVFEEAVKSKAAASSTKEFPDGFWLGEOLV-CWQAGTTPWNIFPVISLYLMGEV 252
Db 282 LLTGPTSPIANIQSDIGASENSD-----GEMVWSCSAISLSDIVF----- 322
Qy 253 TNQSFRTILPOQY-LRPVEDVATSDQDCYK-----FAISQSSTGTVMGAVIMEGFVVVF 306
Db 323 -----TINGVQYPLPSPAYILQSGSCTSGFGMDVPTESGELWILGDVFIQRYETVF 375
Qy 307 DRARKRIGFA 316
Db 376 DRANNQVGLA 385

RESULT 5
B38302
pepsin (EC 3.4.23.-) II-1 precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997
C:Accession: B38302
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucle
A:Reference number: A38302; MUID:91009127
A:Accession: B38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59235; GB:J05638
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.7%; Score 239; DB 2; Length 387;
Best Local Similarity 26.3%; Pred. No. 5.3e-13;
Matches 83; Conservative 56; Mismatches 114; Indels 62; Gaps 14;

Qy 18 SSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPNTVVRANTAAITSD---KFFINGS 74
Db 122 SSTFQATSETLSITYGTGSMTGILGYDTVKV---GNIEDTNOIFGLSKTEPGITFLV-A 176
Qy 75 NWEGILGLAYAEIARDDSPPEFFDSLVKQTHV-PNLFSLQCGAGFPPLNQSEVLASVGG 133
Db 177 PFDGILGLAYPSISASDTP--PVFDNNWNEGLVSEDLFSVYLSNG-----EKGS 224
Qy 134 SMITGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGQDLK--DCKEYNDKSIIVDSG 191
Db 225 MVMFGGIDSSYTGSLNWPVSVGEYQWISVDSITMNGKTIACADSC-----QAVVDTG 278
Qy 192 TTNLRPKKVFEEAVKSKAAASSTKEFPDGFWLGEOLV-CWQAGTTPWNIFPVISLYLMG 250
Db 279 TSLLAGTSAISKTSIQYIGASKNL-----LGENIISCSAIDSLDPIVF----- 321
Qy 251 EVTNQSFRTILPOQYLRPVED-VATSDQDC---YKFAISQSSTGT--VMGAVIMEGFVV 304
Db 322 -----TINNVOYPLPASAYILKEDDDCLSGFGMDVPTESGELWILGDVFIQRYET 372
Qy 305 VFDRARKRIGFAVSA 319
Db 373 VFDRANNQVGLAAA 387
RESULT 6
E38302

pepsin (EC 3.4.23.-) IIis precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 22-Jun-1999
C:Accession: E38302
R:Kageyama, T.; Tanabe, K.; Koizumi, O.
J. Biol. Chem. 265, 17031-17038, 1990
A:Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nu
A:Reference number: A38302; MUID:91009127
A:Accession: E38302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-387 <KAG>
A:Cross-references: GB:M59237; GB:J05639; NID:9165597; PIDN:AAA85370.1; PID:9165598
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.6%; Score 237.5; DB 2; Length 387;
Best Local Similarity 26.0%; Pred. No. 7.2e-13;
Matches 87; Conservative 62; Mismatches 121; Indels 65; Gaps 15;

Qy 2 VPFIYLAQHFLLCSGW-----SSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPNV 54
Db 101 VPSVYCSS--AACSVHQNFPEDSSTFQATSESLITYGTGSMTGILGYDTVKV---GNI 155
Qy 55 TVRANIAAITESDK--FFINGSNWEGILGLAYAEIARDDSPPEFFDSLVKQTHV-PNLFS 112
Db 156 EDNQIFGLSESEPGSFYIYAPFDGILGLAYPSISSDAP--PVFDNNWNEGLVSEDLFS 213
Qy 113 LQLCGAGFPPLNQSEVLASVGGSMIIGGIDHSLTSGSLWYTPIRREWYEVIIIVRVEINGQ 172
Db 214 VYLSDD-----ESGVVWFGGIDSSYTGSLNWPVSVGEYQWISVDSITMDGE 263
Qy 173 DLKM--DCKEYNDKSIIVDSGTTNLRLPKKVFEEAVKSKAAASSTKEFPDGFWLGEOLV 230
Db 264 TIACADSC-----QAIVDTGTSLLAGP---TSAISNIQSYIGASENSD---GEMIVS 309
Qy 231 WQAGTTPWNIFPVISLYLMGEVNTQSFRTILPOQYLRPVEDVATSDQDC---YKFAIS 286
Db 310 CS-----SMTSLPNIV-----FTINGVQYVPVPSAYILEDDACISGFEGMNL 352
Qy 287 QSSTGT--VMGAVIMEGFVVFDRAKRIGFAVSA 319
Db 353 DTYTGELWILGDVFIQRYETVFDRAANNQGLAAAA 387

RESULT 7
I47176
Chymosin (EC 3.4.23.4) precursor - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Jun-2000
C:Accession: I47176
R:Foltmann, B.; Jensen, A.L.; Loenblad, P.; Smidt, E.; Axelsen, N.H.
Comp. Biochem. Physiol. A 68, 9-13, 1981
A:Title: A Developmental Analysis of the Production of Chymosin and Pepsin in Pigs.
A:Reference number: I47176
A:Accession: I47176
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-380 <FOL>
A:Cross-references: EMBL:U14406; NID:9540096; PIDN:AAB08492.1; PID:9540097
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; hydrolase

Query Match 12.5%; Score 235; DB 2; Length 380;
Best Local Similarity 26.0%; Pred. No. 1.2e-12;
Matches 85; Conservative 54; Mismatches 128; Indels 60; Gaps 14;

Qy 2 VPFIY-----LQAHFTLCSGWSSTYRDLRKGVYVPTQGWEGELGTLVSIPIHGPNTV 56
Db 99 VPSVYCSDACQNHHRFNPSKSTFQNLKPLSIQYGTGSGIQGLGYDTVMV--AGIVDA 156

```
QY 57 RANIAAITESDKFFINGSNWEGILGLAYAEIARPDSPPEFFDSLVKQPHV-PNLFSLQL 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 HQTVGLSTQEPDIFTYSEFDGILGLYPELA--SEYTVVPFDNMHRLVAQDLFAVYM 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 CGAGFPLNOSVELASVGGSMITGGIDHSLYTGSWYTPTRRWEYVEIIVRVEINGQDLK 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 S-----RND-----GSMLTIGADPSYTGSLHWVPVMTQLYMQFTVDSVYINGVVA 263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 MD--CKEYNYDKSIIVDSGTTNLRPKKVEAAVKSTKKAASTEKFPDGFWLGE-OLVCWQ 232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 CNGGC-----QALDTGTSMLAGSSDILNTQMAIGATESQ-----YGEFFDIDCS 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 AGTTPWNIPFVLSLYLMEGVNTNQSFRTITLPOQYLPRVEDVATSDQDCY---KFAISOSS 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 LSSMPTVVF-----EISGRMPLP--PSAY-----TNDOGFCTSGFGQDSKS 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 TGTVMGAVIMEGFVYVFDRAKRKIGFA 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 QHWILGVFIQBYYSYVFDRAANNRVGLA 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
S19684
pepsin A (EC 3.4.23.1) 2/3 precursor - Japanese macaque
N:Alternate names: pepsinogen A isozyme 2/3
C:Species: Macaca fuscata (Japanese macaque)
C:Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999
C:Accession: S19684; S16064
R:Kageyama, T.; Tanabe, K.; Koiwai, O.
Eur. J. Biochem. 202, 205-215, 1991
A:Title: Development-dependent expression of isozymogens of monkey pepsinogens and struc
A:Reference number: S19681; MUID:92037645
A:Accession: S19684
A:Molecule type: mRNA
A:Residues: 1-388 <RAG>
A:CROSS-references: EMBL:X59755; NID:g38068; PIDN:CAA42427.1; PID:g38069
A:Note: parts of sequence, including amino ends of pepsinogen and activation intermediat
A:Comment: It could not be determined if this sequence represents isozyme 2 or 3, which
in by 4 months of age.
C:Comment: Although two-step activation is observed, activation is predominantly a one-
C:Superfamily: pepsin
C:Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein diges
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-388/Product: pepsinogen A 2/3 #status experimental
F:16-62/Domain: activation peptide #status experimental <APT>
F:63-388/Product: pepsin A 2/3 #status experimental <ENZ>
F:40-41/cleavage site: Asp-Phe (pepsin) #status experimental
F:62-63/cleavage site: Leu-Ile (pepsin) #status experimental
F:94,277/Active site: Asp #status predicted
F:107-112,268-272,311-344/Disulfide bonds: #status predicted
F:130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 12.5%; Score 234.5; DB 1; Length 388;
Best Local Similarity 26.8%; Pred. No. 1.3e-12;
Matches 88; Conservative 55; Mismatches 128; Indels 57; Gaps 14;

QY 2 VPIYLQIA-----HFTLCGWSSTYRDLRKGVYVPYVYQKWEGLGTLVLSIPHGNVTV 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 VPSVYCSSIACTNHNHNFQDSSSTGSTVSTVYTGTSMTGILGIDYVQV---GGISD 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 RANIAAITESDK--FFINGSNWEGILGLAYAEIARPDSPPEFFDSLVKQTHV-PNLFSLQ 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 TNOIFGLSETEGSLFYAPFDGILGLAYPSIS--SSGATPVFDNLWNOGLYSQDLFSYV 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 LCAGFPLNOSVELASVGGSMITGGIDHSLYTGSWYTPTRRWEYVEIIVRVEINGQDL 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 LSAD-----DQS-----GSMITIGADPSYTGSLHWVPVSVVEGYWQISVDSITMNGEAI 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 KMDCKEYNYDKSIIVDSGTTNLRPKKVEAAVKSTKKAASTEKFPDGFWLGEOLV-CWQA 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 A--CAE--GCQALVDRTGTLTGTPSPITANIOSDIGASENSD-----GEMVWSCSAI 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 234 GTTPWNIPFVLSLYLMEGVNTNQSFRTITLPOQYLPRVEDVATSDQDCYK-----FAISQS 288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 STLQPDIVF-----TINGIQVPVPSAY-----ILQSQGSCISGFGQMDVPTE 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 289 STGTVMGAVIMEGFVYVFDRAKRKIGFA 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 GELWILGDFVFIQYFTVFDRAANNQVGLA 385
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
PEPG
pepsin A (EC 3.4.23.1) precursor - pig
N:Alternate names: pepsinogen A
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Apr-1994 #sequence_revision 22-Jul-1994 #text_change 18-Jun-1999
C:Accession: J03037; A32455; B22434; A91410; A90185; A92039; A92179; PNO145; A00982
R:Tsukagoshi, N.; Ando, Y.; Tomita, Y.; Uchida, R.; Takemura, T.; Sasaki, T.; Yamagata
Gene 65, 285-292, 1988
A:Title: Nucleotide sequence and expression in Escherichia coli of cDNA of swine peps
al protein.
A:Reference number: J03037; MUID:88313677
A:Accession: J03037
A:Molecule type: mRNA
A:Residues: 1-386 <TSU>
A:CROSS-references: GB:M20920; NID:g164601; PIDN:AAA31095.1; PID:g164602
J:Lin, X.; Wong, R.N.S.; Tang, J.
J. Biol. Chem. 264, 4482-4489, 1989
A:Title: Synthesis, purification, and active site mutagenesis of recombinant porcine
A:Reference number: A32455; MUID:89174702
A:Accession: A32455
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-288,290-301,'Y',303-386 <LIN>
A:CROSS-references: GB:J04601; NID:g164603; PIDN:AAA31096.1; PID:g164604
A:Note: replacement of 91-Asp by Ala prevents autocatalytic activation
R:Ichihara, Y.; Sogawa, K.; Takahashi, K.
J. Biochem. 98, 483-492, 1985
A:Title: Isolation of human, swine, and rat prepepsinogens and calf preprochymosin, a
A:Reference number: A22434; MUID:86059312
A:Accession: B22434
A:Molecule type: protein
A:Residues: 1-19,'X','X',22,'X',24-26 <ICH>
R:Moravsek, L.; Kostka, V.
FEBS Lett. 43, 207-211, 1974
A:Title: Complete amino acid sequence of hog pepsin.
A:Reference number: A91410; MUID:74299591
A:Accession: A91410
A:Molecule type: protein
A:Residues: 60-288,290-386 <MOR>
R:Stepanov, V.M.; Baratova, L.A.; Pugacheva, I.B.; Belyanova, L.P.; Revina, L.P.; Tim
Biochem. Biophys. Res. Commun. 54, 1164-1170, 1973
A:Title: N-terminal sequence of swine pepsinogen and pepsin. The site of pepsinogen a
A:Reference number: A90185; MUID:74031413
A:Accession: A90185
A:Molecule type: protein
A:Residues: 16-33,'D',35-118,'SD',121-127,'E',129-134 <STE>
A:Note: the authors point out the similarity of residues 31-43 and 59-71
R:Ong, E.B.; Perlmann, G.E.
J. Biol. Chem. 243, 6104-6109, 1968
A:Title: The amino-terminal sequence of porcine pepsinogen.
A:Reference number: A92039; MUID:69054241
A:Accession: A92039
A:Molecule type: protein
A:Residues: 16-33,'D',35-54,'AE' <ONG>
R:Sepulveda, P.; Marcinszyn, J.; Liu, D.; Tang, J.
J. Biol. Chem. 250, 5082-5088, 1975
A:Title: Primary structure of porcine pepsin. III. Amino acid sequence of a cyanogen
A:Reference number: A92179; MUID:75211282
A:Contents: active site
A:Accession: A92179
A:Molecule type: protein
A:Residues: 58-288,290-322,'D',324-349 <SEP>
A:Note: this is the final paper in a series
```


Matches	90;	Conservative	56;	Mismatches	120;	Indels	67;	Gaps	15;
---------	-----	--------------	-----	------------	------	--------	-----	------	-----

QY	2	VPEIYLQA-----HFTLCGWSSTYRDLRKGVVYPYTOGWEGELGTDLVSIHPGNVTV	56
DB	102	VPSVCSSLACTNHNRFNPEDSSTYQTSETVSITYTCTGSMTGILGYDTVQV---GGISD	158
QY	57	RANIAAITESDK-FFTINGSNWEGILGLAYAEIARPDSPDPFDSLKVQTHV-PNLFSLQ	114
DB	159	TNQIFGLSETPGSFLYAPFDGILGLAYPSS--SSGATPVFDNIWNQGLVSQDLFSVY	216
QY	115	LCCAGFP LNQSEVLASVGGSMIIGGDHSILTYGSLWTPIRRREWWYEVIIVRVINGQDL	174
DB	217	LSAD-----DOS-----GSVVIFGGIDSSYYTGLSNWPVTVEGYWQIIVDSITMNGEAI	266
QY	175	KMDCKEYNVDKSIDVSGTTNLRLPKKVFEEAAVKSIIKAASSTEKFPGDFWLGEQLV-CWQA	233
DB	267	A--CAE-GCOAIVDTGTSLLTGPTSPIANIQSDICASENSD-----GDMVWSCSAI	314
QY	234	GTTPNWIFFVISLYLMGEVTNQSFRTILPQQY-LRPVEDVATSDDCYKFAIS-----	286
DB	315	SSLPDIVF-----TINGVQVPVPPSAYILQSEGSC-----ISGFQGMN	352
QY	287	--QSSTGTVMGAVIMEGFVVVFDRARKRIGFA	316
DB	353	LPTESGELWILGDVFIQRIYTFVDRANNQVGLA	395

RESULT 13

PEMQAR

pepsin A (EC 3.4.23.1) precursor - rhesus macaque

N;Alternate names: pepsinogen A

C;Species: Macaca mulatta (rhesus macaque)

C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Jun-1999

C;Accession: J03039

R;Evers, M.P.J.; Zelle, B.; Bebelman, J.P.; Pronk, J.C.; Mager, W.H.; Planta, R.J.; E

Gene 65, 179-185, 1988

A;Title: Cloning and sequencing of rhesus monkey pepsinogen A cDNA.

A;Reference number: J03039; MUID:88313666

A;Accession: J03039

A;Molecule type: mRNA

A;Residues: 1-388 <ENE>

A;Cross-references: GB:M20788; NID:g342274; PIDN:AAA36902.1; PID:g342275

C;Superfamily: pepsin

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-388/Product: pepsinogen #status predicted <MAT>

F;16-62/Domain: activation peptide #status predicted <APT>

F;63-388/Product: pepsin #status predicted <ENZ>

F;94,277/Active site: Asp #status predicted

Query Match	12.3%;	Score	231.5;	DB 1;	Length	388;			
Best Local Similarity	27.4%;	Pred. No.	2.4e-12;						
Matches	90;	Conservative	54;	Mismatches	126;	Indels	59;	Gaps	14;

QY	2	VPEIYLQA-----HFTLCGWSSTYRDLRKGVVYPYTOGWEGELGTDLVSIHPGNVTV	56
DB	102	VPSVCSSLACTNHNLFNPQDSSTYQTSETVSITYTCTGSMTGILGYDTVQV---GGISD	158
QY	57	RANIAAITESDK-FFTINGSNWEGILGLAYAEIARPDSPDPFDSLKVQTHV-PNLFSLQ	114
DB	159	TNQIFGLSETPGSFLYAPFDGILGLAYPSS--SSGATPVFDNIWNQGLVSQDLFSVY	216
QY	115	LCCAGFP LNQSEVLASVGGSMIIGGDHSILTYGSLWTPIRRREWWYEVIIVRVINGQDL	174
DB	217	LSAD-----DOS-----GSVVIFGGIDSSYYTGLSNWPVTVEGYWQIIVDSITMNGEAI	266
QY	175	KMDCKEYNVDKSIDVSGTTNLRLPKKVFEEAAVKSIIKAASSTEKFPGDFWLGEQLV-CWQA	233
DB	267	A--CAE-GCOAIVDTGTSLLTGPTSPIANIQSDICASENSD-----GEMVWSCSAI	314
QY	234	GTTPNWIFFVISLYLMGEVTNQSFRTILPQQY-LRPVEDVATSDDCYK-----FAISQ	287

Db 315 SSLPDIVF-----TINGVQYPLPPSAYILOSGCTSGFQGMVPT 356

QY 288 SSTGTVMGAVIMEGFVYVFORARKRIGFA 316

Db 357 SGELWILGDVFIROYFTVFDRAANQVGLA 385

RESULT 14

A41443

pepsin (EC 3.4.23.-) precursor, embryonic - chicken

C;Species: Gallus gallus (chicken)

C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 21-Jul-2000

C;Accession: A41443

R;Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.

J. Biochem. 103, 290-296, 1988

A;Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsin

A;Reference number: A41443; MUID:88227903

A;Accession: A41443

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-383 <HAY>

A;Cross-references: GB:D00215; NID:q2760810; PIDN:BARA00153.1; PID:q222853

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 12.3%; Score 230.5; DB 2; Length 383;

Best Local Similarity 24.1%; Pred. No. 2.9e-12;

Matches 77; Conservative 65; Mismatches 115; Indels 63; Gaps 13;

QY 8 QAHFTLCSGWSYRDLRGVYVPYTGKWEGLGTLVSIHPGPNVTVRANIAAITESD 67

Db 113 QSHQMFNPSSQSYKSTGQNLSTHYGTGDMGVGCDTVIVASIMDTNQLFGLST-SEPG 171

QY 68 KFTINSNWEILGLAYAEIARPDSPPEFDSLKVQTHV-PNLFSLQLCGAGFPLNQSE 126

Db 172 QFEVY-VKFDGILGLGYPGLAA--DGITPVFDNNVNESLLEQNLFVYLS----- 218

QY 127 VLASVGSMTIIGDHSLSYLPPIREWYVEIIVRVEINGQDL--KMDCKEYNYD 184

Db 219 -REPMGSMVVFVGIDSEYFTGSINWIPVSQYQWQISMSDIIVNKOEIACSSGC----- 271

QY 185 KSTVDSGTTNLRPLPKVFEAAVKAASSTKFPDGFGLGVLQVQAGTTPWNIPFVI 244

Db 272 QAIDGTSLVAGPASDINDIQAVGANQNT-----YGEYSV-----NCSHIL 314

QY 245 SL-----YLMGEVNTNQSFRIILPQOYLRPVEDVA---TSQDDCYKPAISQSSTGTVMGA 296

Db 315 AMPDVVVFVIGGI-----QY--PVPALAYTEQNGQGTGTCMSSFQNSADLWILGD 360

QY 297 VIMEGFYVVEDRARKRIGFA 316

Db 361 VFIRVYYSIFDRANNRVGLA 380

RESULT 15

C38302

pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997

C;Accession: C38302

R;Kageyama, T.; Tanabe, K.; Koizumi, O.

J. Biol. Chem. 265, 17031-17038, 1990

A;Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequence, and cDNA for embryonic chicken pepsin

A;Reference number: A38302; MUID:91009127

A;Accession: C38302

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-387 <KAG>

A;Cross-references: GB:J05638

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 12.2%; Score 230; DB 2; Length 387;

Best Local Similarity 26.4%; Pred. No. 3.2e-12;

Matches 82; Conservative 52; Mismatches 121; Indels 56; Gaps 11;

QY 18 SSTYRDLRGVYVPYTGKWEGLGTLVSIHPGPNVTVRANIAAITESDKFFINGSWE 77

Db 122 SSTYQGTSETLSITYGTGSMTGILGYDVKVGSIEDTNQIFGLSKTEPFLTF--APFD 179

QY 78 GILGLAYAEIARPDSPPEFDSLKVQTHV-PNLFSLQLCGAGFPLNQSEVLASVGGSMI 136

Db 180 GILGLAYPSISSDAT--PVFDNNWNEGLVSQDLFSYLSSDD-----EKGSLVM 227

QY 137 IGGIDHSLSYTGSLWYTPIRREWYVEIIVRVEINGQDLKM--DCKEYNYDKSIIVDSGTTN 194

Db 228 FGGIDSSYYTGSLNWNVPVSYEGYQWITMDSVINGETIACADSC-----QAIYDVTGTS 281

QY 195 LRLPKKVFEAAVKAASSTKFPDGFGLGVLQV-CWQAGTTPWNIPFVLSLYLMGEVT 253

Db 282 LTGP-----TSAISNIQSYIGASK----NLLGENVISCAIDSLPDIVF----- 321

QY 254 NQSFRTIILPQOYLRPVEDVATSDQDCYKPAISQSSTGT-----VMGAVIMEGFYVVED 307

Db 322 -----TINGIQYPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILGDVFIROYFTVFD 375

QY 308 RARKRIGFAVS 318

Db 376 RANNQLGLAAA 386

Search completed: March 24, 2001, 13:23:12

Job time: 102 sec.


```
RESULT 2
US-09-434-427-2
; Sequence 2, Application US/09434427
; Patent No. 6162630
; GENERAL INFORMATION:
; APPLICANT: POWELL, DAVID J.
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CHAPMAN, CONRAD G.
; APPLICANT: EVANS, JOANNE R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH-70262-D1
; CURRENT APPLICATION NUMBER: US/09/434,427
; CURRENT FILING DATE: 1999-11-04
; EARLIER APPLICATION NUMBER: US 08/999,723
; EARLIER FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: UK 9626022.9
; EARLIER FILING DATE: 1996-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-434-427-2

Query Match 49.3%; Score 925; DB 3; Length 518;
Best Local Similarity 52.3%; Pred. No. 9.9e-95;
Matches 174; Conservative 56; Mismatches 97; Indels 6; Gaps 3;

Qy 3 PRIYLAHFTLCGWSSTYRDLKRGVYVPTQGWEGELGTLVSIPIHGNVTVRANIAA 62
Db 122 PSYIDTYED--TERSTVRSKGFEDVTKYTOGWTGFGVGDVLPKGFNFSFLVNIAT 179
Qy 63 ITESDKFFNGSNWEGILGLAYAEIARDDSPDPFDSLVKQTHVNPVNFSLQCGAGPL 122
Db 180 IFESENFLPGIKWNGILGLAYATLAKPSSLETFDFSLVTOANIPNVFQMCGAGLPV 239
Qy 123 NOSEVLASVGGSMITGGIDHSLYTSLVTPTRRREWYEVIIVRVEINGODLKMCKEYN 182
Db 240 AGS---GTNGGSLVGGIEPSLYKGDWTPTRKEWYQIELKLEIGQSUNLDCREYN 296
Qy 183 YDKSVSDGTTNLRPKPKVFEAAVKSIAKASSTKFPDGFGLGQVLCVQAGTTPWNIFP 242
Db 297 ADKAIVDSGTTLLRLPKQKVDAAVEAVARASLIPFSDGFWTGSOLACWTNSETPWSYFP 356
Qy 243 VLSLYLMGEVNTQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSSCTGVMGAVIMEGF 302
Db 357 KISIYLRDENSRSFRITILPQLYTOPMMAGLNY-ECYRFGISPSNALVIGATVMEGF 415
Qy 303 YVVFDRARRKRGFAVSACHVHDEFTAAVEGPF 335
Db 416 YVIFDRAQRVGFAGFAASPCAEIAGAAVSEISGPF 448

RESULT 3
US-08-208-007A-13
; Sequence 13, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
```

```
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Query Match 12.2%; Score 230; DB 1; Length 396;
Best Local Similarity 25.2%; Pred. No. 3.2e-17;
Matches 84; Conservative 53; Mismatches 134; Indels 62; Gaps 14;

Qy 2 VPFIY-----LQAHTLCGWSSTYRDLKRGVYVPTQGWEGELGTLVSIPIHGNVTV 56
Db 104 VPSVYCTSPACKTHSRFPQSOSSTYSQPSQSFISQYQSGLSGIAGQVSV-EGLTVVG 162
Qy 57 RANIAAITESDKFFNGSNWEGILGLAYAEIARDDSPDPFDSLVKQTHVNPVNFSLQIC 116
Db 163 QQFGEVSTPEPGQTFVD-AEFDGILGLGYPSLA--VGGVTPVFDNMQA-----NLVDLPMF 215
Qy 117 GAGFPLNGSEVLASVGGSMITGGIDHSLYTSLVTPTRRREWYEVIIVRVEINGODLKM 176
Db 216 SVYMSNPE---GGAGSELIFGGYDHSFSGSLNWPVTKQYMWQIALDNTQVGG--TVM 270
Qy 177 DCKEYNDKSVSDGTTNLRPKPKVFEAAVKSIAKASSTKFPDGFGLGQVLCVQAGTTP 236
Db 271 FCSE--GQAIIVDTGCTSLITGSPDKIKQLQNAIGAAP-----VDGEYAVE-----CA 315
Qy 237 PWNIPPVLSLYLMGEVNTQSPRITILPQOYLRPVEDVATSDQDCYKFAISQSSCTG- 291
Db 316 NLNMPDVTFITNG-----VPYTLSPATAY--TLDFVDGMQFQFC-----SSGFQGLD 359
Qy 292 -----TVMGAVIMEGFYVVFDRARRKRGFA 316
Db 360 IHPPAGPLWILGDVFIQFYVDFDRGNRRVGLA 392

RESULT 4
5217891-15
; Patent No. 5217891
; APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.
; TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
; A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,398
; FILING DATE: 09-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,551
; FILING DATE: 28-JUL-1987
; SEQ ID NO:15:
; LENGTH: 458
5217891-15
```

```

US-08-208-007A-12

Query Match      11.9%; Score 224; DB 1; Length 412;
Best Local Similarity 27.2%; Pred. No. 1.6e-16;
Matches 93; Conservative 53; Mismatches 136; Indels 60; Gaps 15;

Qy 2 VPFIYLO-----AHFTLCSGWSTYRDLRKGVVYPYTOGKWGEIGTDLIVSIP----- 49
   ||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 105 VPSIHCKLLDIACWIHHKYNSDKSSYVKNGTSFDIHYGSGSLSGYLSDQDVPVPCQSAS 164
   ||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 50 -----HGPNVTVVRANIAATESDKFINGSNWEGILGLAYAEIARDDSPPEFFDSLKVQT 105
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 165 SASALGGVKVERQFGTEATKGPIFTAARKFDGILGMAYPRIS--VNNVLVFDFNLMOOK 222
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 106 HV-PNLSLOLCGAGFPPLNQSEVLASVGSMIIIGDHSLTGSLWYTPPIRRWNYEVI 164
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 223 LVQNIFSYL-----SRDPDAQPGELMLGTDSKYKGSLSYLVNTRKAYQVHL 274
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 165 VRVEI-NGODLKMCCKEYNDKSVDSGTTNLRLPKKVFEAAVSIKAASTEKFPPDGFW 223
   ||| : | | | | | : ||||| : | | | | | : | | | | | : | | | | | :
Db 275 DQVEVASGLTL---CKE--GCEAVDTGTSLMWGPDVDERELQKAIGAVPLIQ----- 322
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 224 LGEQLV-CWQAGTTPWNIPFYISLYLMEVTNQSFRTILPQQYLRPVEDVATSQDCYK 282
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 323 -GEYNIPEKYST-----LPAILTKLG---KGKLS--PEDYTLKYSQAKTL-CLS 368
   | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 283 FAISO-----SSTCTVMGAVIMEGFVVVDFDRKRKRIGFAVA 319
   : | | | | | : | | | | | : | | | | | : | | | | | :
Db 369 GFMGMDIPPPSGPWLLGDGVFIGRYTYTFVDNRNNRVGFAAEA 410

```

```

Query Match      12.1% Score 227; DB 5; Length 458;
Best Local Similarity 26.0%; Pred. No. 8.7e-17;
Matches      85; Conservative      59; Mismatches 123; Indels      60; Gaps      14;

QY      2 VPFYLQA-----HFTLCGWSSTYRDLRGVVVYPYQGKWEGLGTDLVSIPIHGPNVTV 56
       |||||::| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      177 VPSIYCKSNACKNHRFDQRKSSTFQNLGKDLSLHYTGSMQGLGTYDTVTSNIVDIQQ 236
       ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
QY      57 RANIAATTESDKFFINGNSNEGILGLAYAEIARDDSPPEFFDSLVKOTHV-PNFLFSLQL 115
       ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      237 TVGLSTGEPCGVF-TVAEFGILGMAYPSLA--SEYSIPVDNNMNRHLVAQDLFSVM 292
       ::| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
QY      116 CGAGFPPLNQSEVLASVGSGMIIGIDHSLYTGSWMYTPIRREWYYEVIIVRVEINGODLK 175
       -| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      293 DRNG----QESMLT-----LGAIDPSSYTSLHHWPVTVOQQWQTVDSTISGVVVA 341
       -| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
QY      176 MD-CKEYNYDKSIVDGGTTNLRLPKKVFEEAANKSIKAASSTEKFPDGFWLGE-QLYCWQ 232
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      342 CEGGC-----QAILDPTGTSKLGPSSDIILIOQAICATONQ-----YGEFDIDCDN 387
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
QY      233 AGTTPWNIFPVISLYLMGEVYNOSFRITLPQQOYLRPVEDVATSDDCY---KFAISQS 289
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      388 LSWPTVWF-----EINGKMYPT-PSAY-----TSQDQGFCTSGFQSENHS 428
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
QY      290 TGTVMGAVIMEGYVVFDRARKRIGFA 316
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :
Db      429 QGWILGDVFIREYVSFDRANNLVGLA 455
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | :

RESULT      5
US-08-208-007A-12
; Sequence 12, Application US/08208007A
; Patent No. 5501969
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,007A
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: No. 5501969e
; FILING DATE: No. 5501969e
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN

```


;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-245-756-2

Query Match 10.0%; Score 188; DB 1; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;

QY 18 SSTYRLKGVVYPYTGKWEGLGTLVSIHPGPNVTVRANIAAITEKDFINGSNWE 77
Db 143 SSTYKNGSSFEIRYSGSGMEGVSDVLOI--GDLTPKVDFAEATSEPLAFAGKFD 200
QY 78 GILGLAYAEIARPDSPPEPFDSLKQTHVPLNF---SLQLCGA---GFLPNOSEVLASV 131
Db 201 GILGLAY-----DSISVKNIVPIYKALELDDLPKFAFYLGDTDKDESD 246
QY 132 GSGMIIGGDHSLYTGSLWTPTRREWYEVIIYRVVEINGODLKMCKKEYNDKSIYVDSG 191
Db 247 GGLATGGVDKSYEGKITWLPVRRKAYE-----VSFGVGLGSEYAELOKGTGAADTG 301
QY 192 TTNLRPKKVFEEAAVSIKAASTKFPDGFGLGEQLVCMQAGTTPWNIPPVISLYLMGE 251
Db 302 TSLIALPSGLAEILNAEIGATG-----WSQYAVDCDTRDS---LPDLTLTFAG- 348
QY 252 VTNQSFRTITLPOQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVFDRARK 311
Db 349 -----YNFTITPYDYTLLEVSGSCISAFTPMDFP-EPIGPLAIIGDSFLRKYYSYVDLGKD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 9

US-08-441-750-2
; Sequence 2, Application US/08441750
; Patent No. 5691166
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,750
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/088,633
; FILING DATE: 06-JULY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/678,916
; FILING DATE: 01-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie

;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 9763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-238-0999
;; TELEFAX: 619-238-0062
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 410 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-441-750-2

Query Match 10.0%; Score 188; DB 1; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;

QY 18 SSTYRLKGVVYPYTGKWEGLGTLVSIHPGPNVTVRANIAAITEKDFINGSNWE 77
Db 143 SSTYKNGSSFEIRYSGSGMEGVSDVLOI--GDLTPKVDFAEATSEPLAFAGKFD 200
QY 78 GILGLAYAEIARPDSPPEPFDSLKQTHVPLNF---SLQLCGA---GFLPNOSEVLASV 131
Db 201 GILGLAY-----DSISVKNIVPIYKALELDDLPKFAFYLGDTDKDESD 246
QY 132 GSGMIIGGDHSLYTGSLWTPTRREWYEVIIYRVVEINGODLKMCKKEYNDKSIYVDSG 191
Db 247 GGLATGGVDKSYEGKITWLPVRRKAYE-----VSFGVGLGSEYAELOKGTGAADTG 301
QY 192 TTNLRPKKVFEEAAVSIKAASTKFPDGFGLGEQLVCMQAGTTPWNIPPVISLYLMGE 251
Db 302 TSLIALPSGLAEILNAEIGATG-----WSQYAVDCDTRDS---LPDLTLTFAG- 348
QY 252 VTNQSFRTITLPOQYLRPVEDVATSDDCYKFAISQSTGTVMGAVIMEGFYVVFDRARK 311
Db 349 -----YNFTITPYDYTLLEVSGSCISAFTPMDFP-EPIGPLAIIGDSFLRKYYSYVDLGKD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 10

US-08-441-751-2
; Sequence 2, Application US/08441751
; Patent No. 5831053
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,751
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,756
; FILING DATE: 16-May-1994
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/088,633
;; FILING DATE: 06-JULY-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/678,916
;; FILING DATE: 01-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 9763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-238-0999
;; TELEFAX: 619-238-0062
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 410 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-441-751-2

Query Match 10.0%; Score 188; DB 2; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;
QY 18 SSTYRDLRGVVVPTQGWEGELGTDLVSIPIHGPNTVVRANIAAITESDKFFINGSNWE 77
Db 143 SSTYKNGSSFEIRVSGSMGEGYVSQDLQI--GDLTPKVDFAEATSEPGLAFAFGFD 200
QY 78 GILGLAYAEIARPDSPPEFFDSLKVQTHVPNLF---SLQLCGA---GFPNLQSEVLASV 131
Db 201 GILGLAY-----DSISVNVKIVPPIYKALELDLDEPKFAFYLGDTOKDESD 246
QY 132 GGSMTIGGDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSG 191
Db 247 GGLATFGGVGDKSKYEGKITWLPVRRKAYWE-----VSFDGVLGSEYAELOKTKGAIDTG 301
QY 192 TTNLRPKKVFEEAAVKSIAASSTEFDPGFWLGEOLVCWQAGTTPWNIFPVISLYLME 251
Db 302 TSLIALPSGLAEILNNAEIGATKG-----WSGQYAVDCDTRDS----LPDLTLTFAG- 348
QY 252 VTNQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVWFDRARK 311
Db 349 -----YNFTITPYDVTLEVSGSCISAFPMDFP-BPIGLAIGDSFLRKYYSVYDLGRD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 11
PCT-US92-02521-2
; Sequence 2, Application PC/TUS9202521
; GENERAL INFORMATION:
; APPLICANT: Gleeson, Martin A
; APPLICANT: Howard, Bradley D
; TITLE OF INVENTION: GENES WHICH INFLUENCE PICHIA PROTEOLYTIC
; TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fitch, Even, Tabin & Flannery
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: PCT/US92/02521
;; FILING DATE: 19920321
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/678,916
;; FILING DATE: 01-APR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 50848PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619)552-1311
;; TELEFAX: (619)552-0095
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 410 amino acids
;; TYPE: AMINO ACID
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US92-02521-2

Query Match 10.0%; Score 188; DB 4; Length 410;
Best Local Similarity 24.1%; Pred. No. 1.7e-12;
Matches 74; Conservative 40; Mismatches 147; Indels 46; Gaps 9;
QY 18 SSTYRDLRGVVVPTQGWEGELGTDLVSIPIHGPNTVVRANIAAITESDKFFINGSNWE 77
Db 143 SSTYKNGSSFEIRVSGSMGEGYVSQDLQI--GDLTPKVDFAEATSEPGLAFAFGFD 200
QY 78 GILGLAYAEIARPDSPPEFFDSLKVQTHVPNLF---SLQLCGA---GFPNLQSEVLASV 131
Db 201 GILGLAY-----DSISVNVKIVPPIYKALELDLDEPKFAFYLGDTOKDESD 246
QY 132 GGSMTIGGDHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSG 191
Db 247 GGLATFGGVGDKSKYEGKITWLPVRRKAYWE-----VSFDGVLGSEYAELOKTKGAIDTG 301
QY 192 TTNLRPKKVFEEAAVKSIAASSTEFDPGFWLGEOLVCWQAGTTPWNIFPVISLYLME 251
Db 302 TSLIALPSGLAEILNNAEIGATKG-----WSGQYAVDCDTRDS----LPDLTLTFAG- 348
QY 252 VTNQSFRTILPQQYLRPVEDVATSDCCYKFAISQSSTGTVMGAVIMEGFYVWFDRARK 311
Db 349 -----YNFTITPYDVTLEVSGSCISAFPMDFP-BPIGLAIGDSFLRKYYSVYDLGRD 402
QY 312 RIGFAVS 318
Db 403 AVGLAKS 409

RESULT 12
US-08-360-673-6
; Sequence 6, Application US/08360673
; Patent No. 5679544
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLOYVEROMYCES YEASTS, THEIR
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

APPLICANT: JARA, Patrick
APPLICANT: LEGOUX, Richard
APPLICANT: LOISON, Gerard
APPLICANT: RAZANAMPARANY, Voahangy
TITLE OF INVENTION: Cassette for the expression of an
endothiapepsin precursor in Cryptonectria Parasitica
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.O. BOX 299
STREET: King Street Station, Suite 500, 1800 Diagonal Road
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/115,753
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/758872
FILING DATE: 07-SEP-1991
APPLICATION NUMBER: FR 9011230
FILING DATE: 11-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, BERNHARD D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/332
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
IMMEDIATE SOURCE:
CLONE: preproendothiapepsin
FEATURE:
NAME/KEY: Protein
LOCATION: 90..419
US-08-115-753-2

Query Match 8.3%; Score 155.5; DB 3; Length 419;
Best Local Similarity 25.1%; Pred. No. 7.4e-09;
Matches 82; Conservative 44; Mismatches 130; Indels 71; Gaps 17;

QY 5 IYLOAHFT---LCSG--WSSTYRDLRKGVYVPYTGKWEGLGTLVSIPIHGPNTVRAN 59
DB 147 IYTPKSTTAKLLSGATWSISYGD-----GSSSGDYTTDTVSV--GGLTVTGQA- 194
QY 60 IAAITESDKFFINGSNWEGILGLAYAEIARPDSPFPFDSLKVQTHVFNLFSLQLCGAG 119
DB 195 VESAKVSSFTEDSIDGLGLAFSTL-----NTVSTQOKTFFDNKASLD 242
QY 120 PPLNQSEVLASVGGSMIGIDHSLYTGLSLWTPI--RRWYVEIIVRVEINGQDLKMD 178
DB 243 SPVFTADLGHAPGYTFNGFIDTATYTGITVAVSTKGFWEITSTGYAVSGTFFKSTS 302
QY 179 KEYNKDSIVDSTNLRPKVFEAAVKSIRKAASSTER-----PDPGFWLGEQLVCQA 233
DB 303 ID-----GIADYGTLLYLPLATVVSAYVAQVSGAKSSSSVGGYVFP-----C--S 345

QY 234 GTTPWNIPFVISLYLMGEVNTQSPRITILPOOYLRPVEDVATSQDDCYKFAISOSSTG-- 291
DB 346 ATLPSTFTGV-----GSARI-VIPGDYI-DFGPISGTSSSC--FGGIQSSAGIG 390
QY 292 -TVMCAVIMEGFYVVFDR-RKRIGFA 316
DB 391 INIFGDVAKAAFFVVGATTTLGF 417
RESULT 15
US-08-115-753-33
Sequence 33, Application US/08115753
Patent No. 6017762
GENERAL INFORMATION:
APPLICANT: JARA, Patrick
APPLICANT: LEGOUX, Richard
APPLICANT: LOISON, Gerard
APPLICANT: RAZANAMPARANY, Voahangy
TITLE OF INVENTION: Cassette for the expression of an
endothiapepsin precursor in Cryptonectria Parasitica
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: P.O. BOX 299
STREET: King Street Station, Suite 500, 1800 Diagonal Road
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/115,753
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/758872
FILING DATE: 07-SEP-1991
APPLICATION NUMBER: FR 9011230
FILING DATE: 11-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, BERNHARD D
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/332
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-115-753-33

Query Match 8.3%; Score 155.5; DB 3; Length 419;
Best Local Similarity 25.1%; Pred. No. 7.4e-09;
Matches 82; Conservative 44; Mismatches 130; Indels 71; Gaps 17;
QY 5 IYLOAHFT---LCSG--WSSTYRDLRKGVYVPYTGKWEGLGTLVSIPIHGPNTVRAN 59
DB 147 IYTPKSTTAKLLSGATWSISYGD-----GSSSGDYTTDTVSV--GGLTVTGQA- 194
QY 60 IAAITESDKFFINGSNWEGILGLAYAEIARPDSPFPFDSLKVQTHVFNLFSLQLCGAG 119
DB 195 VESAKVSSFTEDSIDGLGLAFSTL-----NTVSTQOKTFFDNKASLD 242
QY 120 PPLNQSEVLASVGGSMIGIDHSLYTGLSLWTPI--RRWYVEIIVRVEINGQDLKMD 178

Db 243 SPVFTADLGYHAPGYNFGFIDTDTAYTGSITYTAVSTKQGFWEWISTGYAVGSGTFKSTS 302
QY 179 KEYNDKSIYDSGTTNLRPKKVFEEAAYKSIIKAASSTEK-----PDGFWLGEOLVCHQA 233
Db 303 ID-----GIADGTTLTLLYPATVVSAYWQAQVSGAKSSSVGGYVFP-----C--S 345
QY 234 GTPWNIFPVISLYLMGEVTNQSFRTIILPQOYLRFVEDVATSDDCYKFAISQSTG-- 291
Db 346 ATLPSFTFGV-----GSARI-VIPGDYI-DFGPISTGSSSC--FGGIQSSAGIG 390
QY 292 -TVMGAVIMEGEYVVFDRA-RKRIGFA 316
Db 391 INIFGDVALKAAAFVVFENGATTPTLGFA 417

Search completed: March 24, 2001, 13:23:51
Job time: 141 sec

THIS PAGE BLANK (USPTO)